

THE CAUSE OF LOCATION
OF ROADS IN MARYLAND:
A STUDY IN
CARTOGRAPHIC LOGIC

by
David J. Gordon

INSTITUTE OF MATHEMATICAL GEOGRAPHY

**MONOGRAPH SERIES
VOLUME 18**

THE CAUSE OF LOCATION

OF ROADS IN MARYLAND:

A STUDY IN CARTOGRAPHIC LOGIC

DAVID J. GORDON

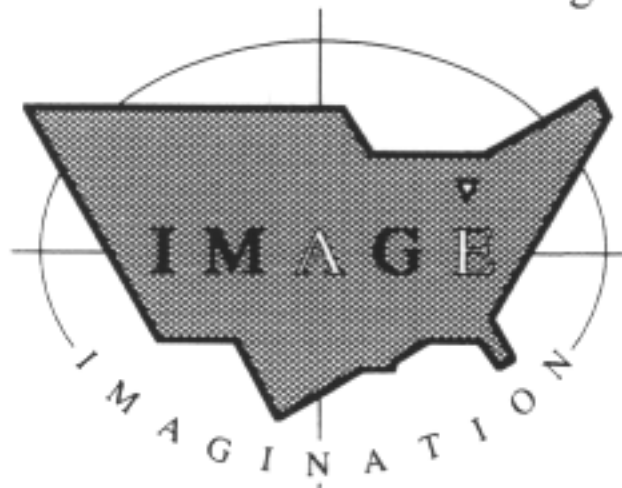


**ANN ARBOR, MICHIGAN
1995**

THE CAUSE OF LOCATION
OF ROADS IN MARYLAND:
A STUDY IN
CARTOGRAPHIC LOGIC

by
David J. Gordon

Institute of Mathematical Geography



© copyright, Institute of Mathematical Geography and David J. Gordon, 1995. All rights reserved, worldwide.

ISBN: 1-877751-57-X

Monographs may be ordered directly from:

Institute of Mathematical Geography
2790 Briarcliff
Ann Arbor, MI 48105-1429

313-761-1231

Dedication:

Edward J. Gordon

1914 - 1975

FOREWORD

This book was originally a thesis submitted to the Temple University Graduate Board in partial fulfillment of the requirements for the degree Master of Arts. The present text is a substantial revision in expression though not in content.

I would like to thank Lane Johnson, my advisor and editor for the original and revision. I would also like to thank Roman Cybriwski and David Cuff for their help, Joan Schiff, Dale Shillito, Joe Revlock, Beverly Agard, Bob Hayes, and Betty Bodine for their kindness, Mark Mattson and the Temple Cartographic Lab for assistance and facilities, and above all my mother for her infinite patience and support.

TABLE OF CONTENTS

Chapter 1 : Introduction to the Cause of Location of Roads	1
Roads	2
The Location of Roads	2
Road Location in Geographic Literature	3
Ridges as Natural Routes	6
A Model of Ridge Routes	7
Routes, Paths, Trips, and Areas	9
Chapter 2 : Developing the Hypothesis of Ridge Roads	12
The Cost of Travel	13
Early Roads	14
Humid Regions of Slight Relief	16
Direct Routes	17
The Benefit a Road Creates	19
Roads And Cities	20
Site Theory	21
Situation Theory	22
Other Sources of Traffic	23
The Area Near a Road	25
Variations in Access Range	27
The Effect of Distance Decay	28
Other Benefit Factors	29
The Hypothesis of Ridge Roads	30

Chapter 3 : Constructing a Model of Ridge Roads	32
Cartographic Logic	32
Sets and Functions	33
Roads Described by Sets and Functions	34
Graphic Construction of Contours	36
A Graph of Travel Cost	37
Routes as Sets of Paths	39
Components of Travel Cost	40
Solving the Distance Resistance Trade-off	41
Paths of Simple Shape	42
A Graph of Crossing Cost	43
A Graph of Detour Cost	44
The Least Cost Overland Path Found	45
The Limit of Detour	46
A Direct Path Defined	46
The Rhombus of Direct Paths	48
Routes as Sets of Trips	49
The Problem of Dimensionality	51
One-to-One Correspondence	51
Location Co-Ordinates	53
Orientation as a Location Co-ordinate	53
The Set of Trips in One Orientation	54
A Boundary That Describes Trips	55
A Boundary That Describes Direct Paths	57
Routes as Bounded Areas	59
A Boundary That Finds Ridge Routes	60
A Contour Map of Travel Cost	62
Road Value Number	63
Confluence Curves	63
Areas That Contain Routes	64
All Routes Described	65
A Model of Ridge Routes Completed	66

Chapter 4 : Central Maryland: a Case Study	71
A Map of Early Roads	72
Roads Correctly Predicted	73
Roads Not Predicted	74
Roads Affected by Other Terrain Features	74
Roads Showing Variation in Detour	75
Roads Affected by Interconnection	76
Other Roads Not Predicted	79
False Predictions	79
Sources of Error	81
Possible Errors in the Map of Ridge Routes	81
Possible Errors in the Map of Early Roads	83
The Cause of Location of Roads Defined	84
Chapter 5 : Generalizing the Results	85
Limiting Assumptions	85
Rivers and Slight Relief	86
Natural Routes and Primitive Roads	88
Land Cost and Early Roads	90
General Assumptions	91
Economic Routes	91
The Endless Road	92
The Limit of Detour	94
Natural Routes and Geographic Law	95
Morphological Law	96
A Single Fact	98
References Cited	99
Appendix A Glossary	102
Appendix B Constructing the Contour Map of Travel Cost	107

LIST OF MAPS

Map One: Ridge Routes and Early Roads in Central Maryland

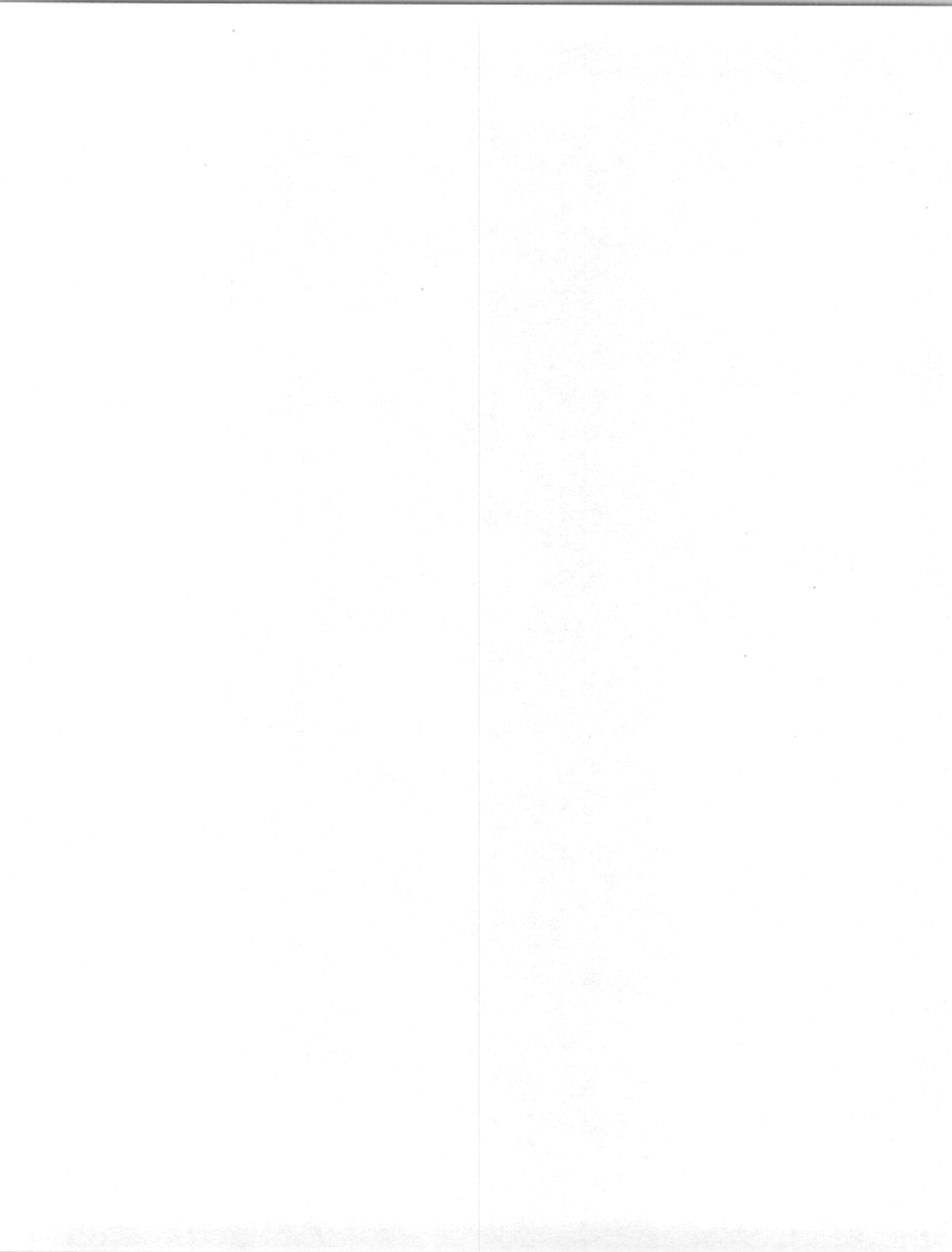
Sheet One	page 68
Sheet Two	page 69
Sheet Three	page 70

LIST OF TABLES

Summary of Roads Predicted and Not Predicted	page 77
--	---------

LIST OF FIGURES

Figure No.	Page	Figure No.	Page
1	33	21	49
2	35	22	50
3	36	23	53
4	37	24	55
5	38	25	56
6	38	26	57
7	39	27	58
8	40	28	58
9	41	29	59
10	42	30	61
11	43	31	61
12	43	32	61
13	44	33	62
14	44	34	63
15	45	35	64
16	45	36	64
17	46	37	107
18	47	38	108
19	48	39	109
20	48		



Chapter 1

Introduction to the Cause of Location of Roads

In plains and lowlands highways may run in any direction expediency suggests, but in mountain areas the pass points the road.

*Ellen Churchill Semple*¹

When Semple says that passes determine the location of roads in mountain areas, she states a generally accepted geographic law. When she suggests that in other areas there is no causal connection between roads and features of terrain, she expresses a widely held opinion. The science of Geography can explain the location of mountain roads, but has no equivalent law to describe the great majority of the world's roads: those outside mountain areas.

"A geographic fact is any fact that can be mapped on the earth's surface, any phenomenon that has an address."² Thus, roads are geographic facts. A road can be mapped, but to say that it *has* an address strains the metaphor. A road *is* an address. The realm of human activity is, in an important sense, not the whole surface of the earth, but the much smaller part of that surface contained in the network of roads. The distribution of people and activity conceived as a static condition is only a time-exposure of a process of distributing occurring on roads. As Cooley said, "Whatever is connected with territorial conditions, with the surface of the earth considered as an area, is connected with

¹ Ellen Churchill Semple, *Influences of Geographic Environment; On the basis of Ratzel's System of Anthropogeography*, (New York: Holt, 1911) p. 548.

² Jean B. Mitchell, *Historical Geography*, (London: The English Universities Press, 1954), p 7.

2. *The Cause of Location of Roads in Maryland*

transportation.”³ Thus roads are central to the subject matter of Geography. What Schaefer called “The science concerned with the formulation of the laws governing the spatial distribution of certain features on the surface of the earth”⁴ should be concerned to formulate the laws that govern the distribution of roads. This paper will present such a law. We will show that in at least one region, the ridge points the road.

Roads

“Roads” in the most general sense includes all persistent paths of travel. Airways and seaways are roads in this broad sense. Constant use makes them permanent features, in effect. But roads that are static, visible objects are confined to the surface of solid land. In this second narrower sense, railroads and canals are roads too, but we will be concerned with only the narrowest sense of the word. One dictionary says: “Road, n, A strip of land along which one may pass from where it is too tiresome to be to where it is futile to go.”⁵ The important part of this definition is that a road is a strip of land. Railroads, Roman roads, and modern highways also have a structure on top of the land, but for the most part, any road before the modern era was a primitive road made of nothing but the natural soil. These early roads are what Brunhes called “basic facts,” that are “the result of a real cooperation between the facts of the terrestrial world and human activity” that he calls on us to “observe first.”⁶

The Location of Roads

A road that is a permanent surface feature has a **location**⁷ which is first its position in space or its geographic coordinates. The position of a road defines what land is under it and what other features it is near. Secondly, the relative positions of its parts gives a road a map shape

³ C. H. Cooley, “The Theory of Transportation,” *American Economic Assoc.*, IX, 3 (May 1894), in Michael E. Eliot Hurst, ed. *Transportation Geography: Comments and Readings* (New York: McGraw Hill, 1974), p. 22.

⁴ Fred K. Schaefer, “Exceptionalism in Geography: A Methodological Examination,” *Annals Assn. of Amer. Geographers* (Sept. 1953) p. 227.

⁵ Ambrose Bierce, *The Enlarged Devil's Dictionary*, Ed. Ernest J. Hopkins, (New York: Doubleday, 1967), p.250.

⁶ Jean Brunhes, *Human Geography: An Attempt at a Positive Classification, Principles and Examples*, trans. I. C. Lecompte, (Chicago: Rand McNally, 1920), pp. 36, 75.

⁷ Terms introduced in **bold face** are defined in the Glossary (Appendix A).

or **alignment**. A road is a long narrow continuous area, so narrow that it can appropriately be described by a mathematical curve without width. The line that describes a road is its **route**. A route is rarely a straight line, and may occasionally take a complex shape. A route can be defined precisely. The edges of a road are sharp, and, in one sense, a road is a small feature, hardly wider than a person, though in the other sense, it approaches a global scale.

The location of a road also includes its relations to other roads. The **density** of roads can be described as the mileage of roads per unit of area, or as the typical distance between parallel roads. **Network structure** describes such characteristics as the number of roads at each junction. This paper will not be concerned with the structure and density of networks. The location of a road can be described without addressing the relations among different roads, if location is understood to mean only the position and alignment of individual routes.

The location of a road is an observable fact, but it is also an action. People locate roads when they choose a strip of land to travel on and to build on. This action can be described by a law of cause and effect. At least in the case of mountain roads, there is a physical connection between travellers and land that makes travel more difficult where the land rises. When this physical work is felt as an economic cost, road builders choose the lowest places. Though the economic connection is less certain in its operation than the physical, the two provide a mechanism through which a pass can act on a road. The results are observably consistent and the road can not act on the pass. With connection, consistency, and independence, it can be inferred that the location of the pass causes the location of the road. A description of cause and effect in a form that allows prediction is a natural law. If it explains geographic facts, it is a geographic law.

Road Location in Geographic Literature

Geography has one law to explain the cause of location of roads. This law is general. It takes many disparate facts, the Brenner, the Karakoram, and the Uspallata, and makes them one fact, *The Pass*. One statement of law replaces many statements of fact, but it describes only a special case. There is no general theory that covers the great majority of the world's roads, those in low hills, plateaux or "plains and lowlands."⁸

⁸ Semple, p. 548.

4 *The Cause of Location of Roads in Maryland*

Since Herodotus,⁹ Geographers have described roads. For Ibn Hauqal¹⁰ in the tenth century, roads provided not just subject matter, but the literary form of Baedeker itineraries, and even a map projection. He draws the roads as straight lines and arranges the rivers and mountains around them. This tradition is what has been called "mere description."¹¹ "Mere" is not a pejorative when used in its root sense of pure and undiluted. Mere description will always do more justice to the variety and complexity of the world than abstract analysis can, but it explains nothing.

In more recent descriptive geography, roads are less central. Hartshorne ignores them, apparently because they are not "areal."¹² A large field of Transportation Geography has developed three analytic approaches, none of which is concerned with the cause of location of roads. The flow approach turns the qualitative distinction that Herodotus¹³ made between the royal road and the mountain track into a quantitative distinction, but it remains description. The graph approach, following Kansky,¹⁴ and the matrix approach, also called the standard model, both depend on simplifying a road network into a topological graph. Cities and other junctions are identified as nodes or vertices, while routes between them are called links or edges. The links of a topological graph have no alignment. As Garrison and Marble put it, "The graph is completely specified . . . as soon as it is known which edges are incident upon which vertices."¹⁵ The complete relationship of a sinuous road to the land under and around it cannot be described by the two points that determine a straight line. The geographic relationship between roads and land is a topographical and not a topological relationship.

⁹ Herodotus, *The Histories*, Trans. Aubrey de Selincourt, (Harmondsworth: Penguin, 1972).

¹⁰ Ibn Hauqal, *Configuration de la Terre (Kitab Surat al-Ard)*, Introduction et Traduction par J. H. Kramers et G. Wiet, (Paris: Editions G. P. Maisonneuve et Larose, 1964).

¹¹ Schaeffer, p. 228.

¹² Richard Hartshorne, "The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past," *Annals Assn. of Amer. Geographers*, XXIX, 3 Sept. 1939.

¹³ Herodotus, *op. cit.*

¹⁴ Karol Kansky, *Structure of Transport Networks, Relationships Between Network Geometry and Regional Characteristics*, Dept. of Geography Research Paper No. 84 (Chicago: Univ of Chicago, 1963),

¹⁵ W. L. Garrison and D. F. Marble, "Graph Theoretic Concepts" reprinted from *A Prolegomenon to the Forecasting of Transportation Development Research Reports*, Transportation Center, Northwestern Univ., Evanston: Northwestern, 1956, pp 40-68 and 89-96, in Eliot Hurst, p. 62.

Taking the word in its broad sense, the most complete analysis of road location as a topographical problem is *The Economic Theory of the Location of Railways* by A. M. Wellington,¹⁶ a civil engineer. Wellington analyzes the physical and economic connection between geographic conditions and the choice of route to extract design principles, but he applies them only in a trial and error method, and speaks of “an art, as distinguished from a science.”¹⁷

Geographers who discuss the cause of location of roads are few. The explanation they most frequently offer is that roads connect cities or other settlements. Kolars and Malin say, “The location of railroads in Turkey reflects relative situational advantage with respect to major population concentrations.”¹⁸ and J. S. Wood says of New England, “Once the site for a meeting house had been set, . . . roads were laid out to it.”¹⁹ Since cities are vertices, the topological graph approach also contains implicit this common idea that roads connect cities.²⁰

There is a connection and a consistent correlation between roads and population centers, but it is not clear in which direction the connection operates. It is widely accepted that cities rise on roads. Thus the common view that roads are determined by settlement patterns contains a logical fallacy. A cause must be independent of its effect and the independence of cities from roads has not been demonstrated.

The independence of terrain needs no demonstration, yet it tends to be treated as a minor factor, operating only under the special conditions where roads follow passes and valleys. Kolars and Malin consider only “major topographic features, coasts [and] impenetrable mountain chains with such natural passes as the Cilician Gate.”²¹ Taafe, Morrill, and Gould discount terrain. They say “Much of the impact of [terrain] on the transportation system is

¹⁶ Arthur Mellen Wellington, *The Economic Theory of the Location of Railways; An Analysis of the Conditions Controlling the Laying out of Railways to Effect the Most Judicious Expenditure of Capital*, (1887; New York: John Wiley and Sons, 1899).

¹⁷ Wellington, p. 831.

¹⁸ J. F. Kolars and H. J. Malin, “Population and Accessibility: An analysis of Turkish Railroads,” *Geographical Review* Vol. 60 (1970), pp. 229-246, in Eliot Hurst, p. 112.

¹⁹ J. S. Wood, “Elaboration of a Settlement system: The New England Village in the Federal Period,” *Jour. of Hist. Geography*, CIV (1984) p. 334.

²⁰ Wellington, too, mentions “cities, towns and other sources of traffic” (p. 51) as a consideration, though he gives them little weight.

²¹ Kolars and Malin, p. 116.

expressed through [its] relationship to the population pattern,"²² which seems to suggest that terrain affects roads only because it affects population density. That terrain is insignificant is also implicit in any approach that tries to predict road location on a uniform plane. For instance, Beckmann says, "Let us assume the absence of any topographic features."²³

Passes are specific features of terrain that are the best places for travellers to pass through. They are **natural routes**. River valleys may also be places where travel is easy and in regions of high relief these two forms of natural route all but completely determine the roads.

Outside mountains, there is no obvious relation between terrain and roads, but it does not therefore follow that there are no natural routes. Where features of terrain are subtle, it may take subtle observation to identify them. Whether natural routes exist outside mountain areas remains an unanswered question. Important as it is, this question receives remarkably little consideration in the academic literature.

Ridges as Natural Routes

Just a few geographers have noticed cases of roads that follow ridges. Pierce Lewis, describing New Orleans, says "a road had been run along the top of Gentilly Ridge."²⁴ A. E. Smailes mentions "routes running West East along the belts of dry, relatively open country afforded by the chalk ridges among the forest-choked lowlands."²⁵ And Ullman has mentioned "Route sixty-six running along the broad ridge between the Meramec and Bourbeuse Rivers."²⁶

Other cases of ridge roads can be found. The Great North Road leaving London through Barnet follows the watershed of the Lea and the Brent. The ancient port of Olbia and the modern port of Odessa are at the beginning of a road on a watershed ridge that reaches some

²² E. J. Taaffe, R. L. Morrill and P. R. Gould, "Transport Expansion in Underdeveloped Countries: A Comparative Analysis," *Geographical Review*, vol. 53 (1963) pp. 503-529, in Eliot Hurst, p. 395.

²³ Martin Beckmann, "Principles of Optimum Location for Transportation Networks," in *Quantitative Geography, Part 1 Economic and Cultural Topics*, Northwestern Univ. Dept. of Geography (Evanston, 1967), p. 97.

²⁴ Pierce Lewis, "New Orleans, The Making of an Urban Landscape," in John S. Adams, Ed. *Contemporary Metropolitan America* (Cambridge: Ballinger Publishing Co., 1976) p. 136.

²⁵ A. E. Smailes, *The Geography of Towns*, (London: Hutchinson, 1953), p. 41.

²⁶ Edward L. Ullman, *Geography as Spatial Interaction*, Ed. Ronald R. Boyce, (Seattle: Univ. of Washington Press, 1980), p. 151.

three hundred miles toward L'vov. This recurring pattern of roads that match watershed ridges suggests a causal connection, but it appears that no geographer has proposed that there might be a general law.

A Model of Ridge Routes

That the pass points the road stands as a geographic law because it is easy to demonstrate the consistency with which it operates. Spotting passes is a simple predictive model that identifies cases with the presumed cause so they can be compared to cases with the effect. In other sciences, these cases may be conceived in terms of time, as that the cause is present *when* the effect is, but geographic facts are differentiated in space. The cause of a geographic fact must be present *where* the effect is. It must itself be a difference between locations, itself a geographic fact. Geographic law describes connections between spatial distributions.

No quantitative model will serve to describe these connections. The question "where?" is not answered by the statement "how much." Only a map can define with precision the curving, continuous shape of a route. A prediction of where roads would be if they followed natural routes is a map of natural routes. If this map matches a map of roads, it will demonstrate the consistency without which there can be no natural law.

A map of roads can be based on the directly observable surface features used by travellers. A map of natural routes must be logically derived. A model must reach its prediction by strict implication. When the hypothesis is assumed true and conditions are given, the prediction should follow as a necessary logical conclusion. Where both the conditions and the effects are spatial patterns that are easily and naturally represented in the map, the logical connection between them should also be represented in the map.

Abler, Adams and Gould have said:

Although maps can summarize much, there are aspects of networks and flows they cannot effectively cope with. The major limitation is their computational possibilities. A map is far too concrete and detailed to permit much in the way of direct manipulation and analysis.²⁷

But this statement is not true. A map is a fundamental logical system. Like the Venn diagram, it represents logical categories with bounded areas. Relations between categories can also be

²⁷ Ronald Abler, John S. Adams and Peter Gould, *Spatial Organization; The Geographer's View of the World*, (Englewood Cliffs: Prentice Hall, 1971), p. 255.

8 *The Cause of Location of Roads in Maryland*

defined this way. In his classification of climates, Köppen²⁸ constructs formal logical arguments of considerable complexity using the logic of overlapping boundaries.

By analogy to a truth table, superimposing two maps can be a formal proof of the biconditional "cause if and only if effect," but to prove a universal generalization by truth tables is possible only if all cases can be defined. This is impractical for most sciences and, to overcome this limit, they use controlled conditions or statistical sampling. Geography does not have the same problem. All possible cases of geographic location are contained in the Earth's surface. When represented on a map, this can be completely described and classified and systematically exhausted.

Reading passes from a contour map is a simple example of the logic of the map. Because, in high relief, elevation dominates the cost of travel, a contour of elevation describes the working connection through which terrain acts on roads. It is, in effect, a contour of relative travel cost. Travel is more difficult on the high side, so travellers stay on the low. The contours surround the peaks and the space between them locates the natural route.

A similar method can locate natural routes in regions of slight relief. Here, there are more variables in the travel cost equation, but the method Thünen²⁹ uses to derive a map of land use areas can be adapted to it. Thünen constructs contours by graphing an economic equation and geometrically projecting the graph into a boundary. In a sense, Thünen's boundaries are contours of travel cost. Similarly constructed boundaries can find ridge routes, much the way contours of elevation find passes. But two problems that are deceptively easy to solve for a model of passes no longer have simple solutions without the mountain range. First, crossing the range neatly defines the goal of a road, or the benefit side of the economic equation. Second, the line of the range puts all possible cases into a more or less neat and regular order.

²⁸ Clyde P. Patton, Charles S. Alexander and Fritz L. Kramer, Chapter 19, *Physical Geography* (Belmont, Cal.: Wadsworth, 1970).

²⁹ J. H. von Thünen, *Der Isolierte Staat in Beziehung auf Landwirthschaft und Nationaloekonomie* (1826; Berlin: Wiegandt, Hempel & Parey, 1875).

Routes, Paths, Trips, and Areas

Most of the tools needed to derive a map of ridge routes from a map of rivers can be found in the geographic literature. One that cannot is an adequate logical conception of the nature of a route. It is a conventional idea that a road goes from point to point, from city to city, or to other identifiable places. However useful it may sometimes be, this concept does not really describe roads. Roads pass through towns, and they do not end at junctions. A crossroads is where two roads cross, not where four roads terminate. Large junctions may be easier to rationalize as points with radii, but they can also be seen as the overlapping of multiple forks and crossroads in which every route has several continuations.

Roads do not even end at the ocean. As R. S. Platt said of the Central American railroads, "From the traffic point of view these railways are not detached units; they are ramifications of the main highway system, which is the sea."³⁰ Passing through ports and junctions, roads can circle the earth and keep going. Though cases do occur where a road comes to a definite end, these should be seen as the exceptions. Travellers go from point to point but roads extend indefinitely.

Since a road does not end, a route, the line that describes a road, has no endpoints. Thus it differs from a path. When the words path and route are loosely used, they may be interchangeable, but there is an important difference. In a strict sense, a **path** is the sequence of locations of a moving object, in this case a traveller. When the traveller comes to a rest, his path has an endpoint. A road does not come to a rest. It accumulates the paths of many travellers. As some leave, others enter. A route and a path are each a continuous curve with complex alignment, but a path has endpoints and a route does not.³¹

The endpoints and alignment of a path define the distance travelled and the surface passed over. The work required to make a trip can be known only when a path has been defined. Travel cost is undefinable on a route without definite length. If the route is cut

³⁰ Robert S. Platt, "Central American Railways and the Pan-American Route," *Annals Assn. of Amer. Geographers*, XVI no. 3, (Mar. 1926) p. 16.

³¹ Paths are directional, and so too, sometimes, are routes, but this is a complication that can safely be ignored. We will refer to the **origin** and **destination** of a path to mean one endpoint and the other without implying a direction of travel.

into segments, the segments correspond to paths. Cost is definable for segments of routes, but the benefit created by a road depends largely on conditions beyond the ends of any route segment. It depends not on where the road lies, but on where it leads. Though the problem has many ramifications, we will show that endless extension alone is a valid definition of the goal of a road.

The two sides of the economic equation can be defined separately, using the concept of the endless route for benefit, and using the concept of a collection of paths for cost. The logic of bounded areas introduces a third distinct concept. Boundaries describe areas. Mapping a road means constructing boundaries to an area that contains the route. We will call this a **route area** to distinguish it from the route it contains. Since the area that contains an endless route must extend across the map, this characteristic of a natural route can be judged from a route area.

A system of boundaries can classify and describe every point on the map. Boundaries provide a means to logically select from all possible locations. Unfortunately they describe a difference between points and no point has the characteristics of extension and alignment that are essential to a route. Though the map contains all places, all possible locations for roads includes not just all places, but all shapes of route.

The set of all paths can define all possible routes. Since every route segment coincides with some path, a route is a set of paths. Paths have extension and alignment so it is possible to select paths that have characteristics of natural routes. The set of all paths could be defined as all curves between endpoints, but short paths of simple geometric shape can closely approximate any realistic shape of route. From this simpler set we will find it possible to select exactly one path for any pair of points, the only one that could be a segment of a natural route.

To associate paths with map areas requires yet a fourth conception of a route. Paths can be associated with points on the map through their endpoints. Since route segments can be cut anywhere, all points in a route are endpoints of route segments. A route is a set of pairs of points. They are the pairs of points that are the endpoints of the paths that are segments of the route. This admittedly elaborate construction associates the characteristics of natural routes to areas on the map. With one path found for any pair of points, it becomes possible to associate travel cost to pairs of points. A natural route is a set of pairs of points with low travel cost.

For reasons by now clear, we will use the term **trip** to mean a pair of points that are the endpoints of paths. Relating pairs of points to single points raises one more difficulty in a problem of dimensionality. This we will resolve by examining trips in one orientation at a time.

By using in all four different conceptions of a route, the endless curve, a set of paths, a set of trips, and an area between boundaries, a correspondence can be established between travel cost and points on the map. With this connection made, contours of relative travel cost can be constructed using an adaptation of Thünen's method. A graph of travel cost can be projected into a boundary such that travel is easier on one side. A contour map of travel cost can be constructed using Euclid's tools on a map of rivers.

This logically derived map will identify all and only those locations that have the characteristics of natural routes. It predicts where early roads should be. We will apply this model to one region, central Maryland between Chesapeake Bay and the mountains. We will find that in this region of slight relief, early roads follow ridges.

Chapter 2

Developing the Hypothesis of Ridge Roads

Ridge Road runs south from Cleveland and Ridge Road leaves New Haven going north. Blue Ridge Road is southwest of Houston and Ridge Road passes east of Cheyenne. One of America's most common place names relates ridges to roads. Ridge Pike enters Philadelphia running exactly on top of the drainage divide between Wissahickon Creek and the Schuylkill River. The ground falls sharply away on both sides. The top of the ridge is a difference between locations at a small enough scale to distinguish where the road is from where it is not. The ridge is the product of ancient uplift and erosion and is unquestionably independent of the road. There is something here to suggest a relationship of cause and effect, but a cause must also be consistent and it is clear that not all roads follow ridges.

On the modern map, Philadelphia's Ridge Pike is embedded in a network of streets that show a great variety of relations to terrain. One snakes along a valley floor, another leaps from hilltop to hilltop on a series of bridges, a third follows a rigid straight line that plunges off a cliff. We can not expect a single cause to explain them all.

Early maps show a simpler pattern. For instance on the Scull and Heap map of 1777,³² only four roads enter Philadelphia from the northwest. Each one runs on one of the four main watersheds. In other directions the relationships cannot be as neat, but all the early roads at Philadelphia show some preference for ridges. Among the roads mapped in the

³² Scull and Heap, "Map of Philadelphia" Engraved by William Faden London (1777) Reproduced Ithaca: Historic Urban Plans, 1969.

late eighteenth century, there is some appearance of consistency. Demonstrating this consistency will be the major question for this paper, but first it is necessary to examine whether there is a working connection, in some sense mechanical, through which the ridge can act on the road with sufficient power to override other possible causes.

The Cost of Travel

On any road, the physical work of changing the location of an object is a large part of the economic cost. The laws of simple mechanics define work of moving a load, or **haulage cost**. Whether the object moved is the traveller himself or a shipment of goods is a distinction that seldom makes a difference. The traveller applies force to the land and the equal and opposite reaction causes him to move. That force times distance equals work remains true whether the force is applied directly through the traveller's feet or through an economic and mechanical transfer as complex as a railroad.

The work required to change the elevation of an object equals mass times rise. The rise of a road is usually expressed by **gradient**, which equals rise per distance. The elevation of a road is almost always close to that of the land it is on, but its gradient is affected by the angle between the road and the hill, and so on the alignment of the route. The slope of the hillside is a maximum limit to the gradient of the road.

Where differences of elevation are extreme, they describe practically the whole cost of travel but other factors do occur in the laws of basic mechanics. Of these mass and acceleration are not differences between locations and cannot cause location. Distance is of special importance, and we will return to it later. Friction remains.

The force applied to moving an object must exceed the force of friction which equals normal force times the coefficient of friction. Normal force can be treated as weight, and therefore as a constant. The coefficient of friction is empirically defined by measuring force and distance. Friction is whatever absorbs work without converting it into motion. The coefficient of friction is normally constant for any given mechanical system. In the mechanical system of a vehicle on a road, part of the friction is attributable to the vehicle and is not a geographic difference. Part is in the surface of the road. Soft surfaces deform and absorb work without converting it into motion, so the traveller's work is wasted. Rough surfaces also waste work. The work of travel is equal to distance times friction

added to distance times gradient. If the road is level, travel cost is least when the road surface is hard and smooth.

Friction can vary from place to place but may not be connected to the friction on the road. Modern road builders can construct a hard, smooth surface on top of the land. Work done reducing friction and gradient is the **capital cost** of the road. It can reduce the cost of travel, but it is limited by traffic volume. Economically, the saving to each traveller must exceed his share of the investment. At one extreme of high traffic and high cost, the Pulaski Skyway flies over the marshes of Newark Bay on a series of spans between pylons and the surface of the road is entirely independent of the surface of the land. At the other extreme, where traffic is thin and capital investment small, roads are closely dependent on natural conditions.

Early Roads

The roads that were mapped in America soon after the revolution represent a minimal capital cost. As Seymour Dunbar³³ has described them, "A so-called wagon road . . . was usually a narrow, winding trail across the country made of nothing but the natural soil."³⁴ When the Europeans brought wagons to America, they took them into the woods. Dunbar says, "The chief obstacles to be overcome were the undergrowth in the forest itself, or dense thickets on lands that held no large timber."³⁵ But Cooper, an observer closer in time, suggests, "The American forest admits the passage of horse, there being little underbrush and few tangled brakes."³⁶ Even today fallen trees are a greater obstacle to travel in the woods than standing ones are and it can only have been worse in a forest that had never been harvested. Whatever the case exactly, an axe was a necessary tool for a traveller, to clear his own way. If the work of clearing survived to benefit other travellers, it would be theoretically a capital cost, but it was inseparable from the work of travel.

³³ Seymour Dunbar, *The History of Travel in America*, (1915; reprint New York: Greenwood Press, 1968).

³⁴ Dunbar, p. 195.

³⁵ Dunbar, p. 137.

³⁶ James Fennimore Cooper, *The Last of the Mohicans*, (1826; Philadelphia: Porter and Coates, no date), p. 320n.

A strip of land, used for travel, but otherwise little changed from its natural condition, is a **primitive road**. Its surface is identical to the surface of the land. Excluding only the costs of the vehicle, costs that do not vary from place to place, the whole economic cost of such a road is what we will call **overland travel cost**, the work it takes to move a load across the natural surface of the land.

Early European travellers did not enter an empty land. "Practically the whole present day [1912] system of travel and transportation in America East of the Mississippi River, including many turnpikes, is based upon or follows, the system of forest paths established by the Indians hundreds of years ago."³⁷ Dunbar suggests "They had a wonderful faculty for establishing routes that were, in an economic sense, the best that could be chosen,"³⁸ but no such "wonderful faculty" need be predicated. An Indian trail is nothing but the mark left on the forest by repeated travel through it. The fact of its being a trail gives it no practical advantage over any other place in the woods so, if travellers were to find a better route, the trail would no longer be trampled by steady use, the forest would grow back, and the trail disappear. Those who chose the routes were those who did the work, so the economic connection between travel cost and location was close and strong. With time for trial and error, it is only natural that they should follow **economic routes**. Trails were located where they would create the most benefit at the least cost.

Before the wagons there was what Dunbar calls "an age of packhorse travel."³⁹ First horses, then wagons need a wider clearing than walking people do, and each places a heavier load on the surface, but all share a common interest in a smooth, solid, level route. "Trails . . . were gradually being widened without official action and changed to roads by the increasing travel over them, and the governmental purpose was merely to hasten and improve a process that had already begun."⁴⁰

Even as capital investment rises, the trade-off between capital cost and operating cost is not related to a choice between alternative routes. As long as a road is not an elaborate structure, the work required to make it firm, smooth, and level depends primarily on how nearly firm, smooth, and level the land already is before the work begins. Thus, both the

³⁷ Dunbar, p. 19.

³⁸ Dunbar, pp. 18-19.

³⁹ Dunbar, p. 194.

⁴⁰ Dunbar, pp. 29-30.

capital cost and the operating cost of an early road are closely dependent on overland travel cost.

The long slow process of struggle against the land establishes a mechanism through which the land can act on roads. There is a physical connection between the nature of the surface and the work of moving a load. There is an economic connection between the work of travel and the choice of route. Roads that develop gradually out of trails will tend to follow natural routes, where the land itself is firm, smooth, and level in its natural state.

Humid Regions of Slight Relief

In large areas of the world, differences of elevation are too slight to have any large effect on the cost of travel. "A horse can, without difficulty . . . ascend gradients of one in thirty on a macadamized surface without sensible diminution of speed."⁴¹ At this rate, a road can rise five hundred feet in three miles with no appreciable effect on the cost of travel. Regions that lie entirely below five hundred feet are sometimes hundreds of miles wide. In plains and low hills the general trend of elevation can have no economic effect on roads. Occasional local features might have steep slopes but there is no reason to suspect that differences of elevation override all other forces in regions of slight relief. The work of travel is still governed by the laws of mechanics but friction can make a difference. How smooth and solid the land is can vary from place to place, even as its level can.

In humid regions water shapes the land. As running water accumulates in streams, it carves itself an efficient channel with steep banks and a rough bed. It carries off the small particles and leaves a rough surface of larger ones. The whole surface of the land is carved into river valleys with always, necessarily, the steepest slopes down toward the

⁴¹ "Road," *Encyclopaedia Britannica*, 1911, XXIII, p. 389.

The specific gradient of one in thirty refers to a macadamized surface. It is not strictly applicable to a primitive road, which is much less smooth, but it overstates the effect of gradient. Friction and gradient are interrelated through the power to weight ratio of the vehicle. A vehicle with a high power and a low weight climbs and accelerates easily. When the surface is rough, the power to weight ratio must be high to overcome rolling resistance. Though the power of the horse drawn wagon is low, its weight is very low and the power to weight ratio is therefore high. With this relatively high power, hills are easily overcome. At the other extreme, a railroad has a negligible rolling resistance. Though its power is great, its weight can be greater still. The power to weight ratio of a train is therefore very low compared to the horse-drawn wagon. A low power to weight ratio makes a railroad much more sensitive to gradient than is a primitive road. Thirty to one is a very steep railroad gradient. The rougher the road is, the less it is affected by changes in elevation.

stream and, of course, if the stream is big enough, the mass of water itself can stop, or even carry off, the traveller.

Where rivers slow down, they build land out of mud. Deltas and bottomlands are level, but soft. In regions of low relief, these lands are extensive, but slight relief means that the land is above the river, drainage is complete, and river-built lands occur only near the streams. All around the rivers there is soft ground, rough ground and steep slopes, the land on which travel is most difficult.

There is a normal pattern of humid regions of slight relief. The rivers form one pattern that covers the land. The spaces between them form a complementary pattern of ridges. Their sides are cut by tributary creeks, but their tops form a partly connected network of minimal slope. Slight relief is almost synonymous with old erosion surfaces. Rainfall weathers rock and rounds the hilltops. It sustains the dense vegetation that builds the deep organic soils of forest and grasslands. The soil fills in the contours of the land. The gentle erosion by water working its way through a root system smooths it more.

Organic soils make a fine firm surface that will carry any load as long as it stays dry. When wet, it softens and wheels and feet sink in, and the travellers' work is converted into deformation of the surface, digging ruts in the road. The ruts gather water thus beginning a destructive cycle in which the road will never dry out. If a road is in a low place where water flows toward it, and it is a primitive road without any major investment in drainage and grading, then it will decay into an impassable morass. On the other hand, if the road stays on the high ground, water will flow away from the road. Watersheds shed water. The firm smooth and level land on which travel is easy is along the top of the ridge.

Where land conforms to the normal pattern of humid regions of slight relief, the friction of the natural surface is greatest at and near rivers. Those slopes that can affect the cost of travel are in the same distribution. These two conditions can be generalized as **resistance**. All the costs of a primitive road are dependent on overland travel cost, which can be defined as distance times resistance. Resistance is nowhere low, but it is highest at and near the rivers.

Direct Routes

When roads follow ridges, they do not usually match the river basin boundaries. Even the most obvious cases of ridge roads sometimes cross rivers. Ridge Pike crosses the Wissahickon

where that stream enters the Schuylkill and the ridge between the rivers ends. The Great North Road continues across London Bridge. To explain the location of these roads is to explain not just why they are usually on watersheds, but also why, and where, they sometimes cross rivers.

There is one obvious difference between roads and watersheds. Watersheds are circuitous, they surround river basins, but roads must be direct. Distance is perhaps the most important factor in the cost of travel. The largest component of the distance a traveller goes is the length of the straight line between two points. This distance is an absolute minimum that cannot be affected by the location of a route. Since no choice of route can reduce it, it can have no influence on that choice.

The distance following a road can easily be more than the straight line distance. A circuitous route can increase distance indefinitely. The difference between road distance and straight line distance is **detour**. It is the effect of the alignment of the route. If a route is close to straight, distance is close to the minimum. Thus, alignment is a characteristic of the location of a road that has a significant effect on travel cost. When a road bends, distances following it rise, and the cost of travel rises too.

Though detour is least when a route is straight, slight bends increase distance only trivially. This effect can be demonstrated with a taut string tied between two nails. It will be straight, but a very slight pressure will be able to bend it out of line. Pulling it farther gets more difficult, but the first noticeable displacement requires only a minute stretching of the string. Similarly, a route can be diverted to the side with only a trivial increase in distance. As long as it remains **direct**, detour is negligible. Though literally "direct" means straight, we use it here in a relative sense to mean nearly straight.

In a region where rivers are ubiquitous, a straight line cannot avoid rivers but a direct route can. Slight bends that add almost nothing to the length of the route may be enough to take it around the headwater of a river. Thus an increase in distance may result in a reduction in the cost of travel.

There is a trade-off between distance and resistance. If a small bend will avoid a river crossing, distance times resistance will be less on the longer route. If crossing a river can avoid a long detour, travel cost will be less with more resistance and less distance. Finding the route with the lowest overland travel cost means balancing distance against resistance.

Examining the roads on various early maps, it can be seen how they solve this balance. The angles at which roads bend can be measured by approximating their curved shapes with straight line segments. It is rare to find a road with an angle inside the bend that is tighter than 150 degrees. Early roads may twist and turn and could fairly be called winding, but, at the same time that they make frequent bends, each road maintains a constant general orientation within a few degrees either side of a straight line.

Because a road must be direct, a river has a long-range effect. It casts a shadow over the land before and beyond it, an area where a direct route can only lead into the river. A direct route that avoids rivers must also avoid the shadows of rivers. A road that leads into the mud is as bad as one already there. The shadow effect of rivers depends on the orientation of the route. A road parallel to a river may run beside and never cross it. A road perpendicular to a river cannot avoid a crossing if it is to remain direct.

The shadow of a river is the opposite of a natural route. Though all the land not at rivers may be equally easy to travel on, it does not all make equally good roads. The shadow effect of rivers divides this land between areas that are in shadow and are therefore unsuitable for roads, and other areas that are outside the shadows of rivers and may contain natural routes.

Roads that follow the route on which all the costs of a primitive road are least follow watersheds wherever these provide direct routes. But it may be that no such route exists, across the grain of the country or where the watershed is circuitous. It then becomes advantageous to cross some rivers and piece together sections of different watersheds to form direct routes that avoid most rivers, but not all. These are ridge routes, the natural routes of humid regions of slight relief.

The Benefit a Road Creates

Where a road “goes” may be taken to mean either through what places it passes or at what places it arrives. Ridge routes explain only the first. Most explanations of the location of roads are concerned with the second issue, identifying destinations. Specific destinations are geographic conditions that make travel more frequent or more useful at some places than at others. They could act on a road through the benefit side of the economic equation. Terrain acts only through the cost side. Examination of the benefits created by a road will show that destinations can have little or no effect on early roads.

Roads And Cities

The idea that roads go to specific destinations most often takes the form that roads connect cities and other settlements. Because cities are concentrations of population, they are places more travellers originate. Because they are concentrations of commercial and social activity, they are places more travellers seek to go. Even though they define only a few points and not a continuous route, cities are small enough to narrowly define the positions of routes. Cities and other settlements are geographic conditions at an appropriate scale that are economically connected to roads.

They are also consistently related. Brunhes says:

The concentration of habitations keeps pace with the concentration of paths of communication. The larger the city, the finer the network of roads which surround it. Inversely, the more physical conditions favor the concentration of roads at one point, the more possibilities of growth a city has.⁴²

But Brunhes describes a two-way connection. Roads go to cities, but also cities rise on roads. A causal connection can not be inferred unless the cause is independent of the effect, and there is little reason to suspect that the locations of cities are independent of the locations of roads.

Smailes says:

Offering certain situations the quality of natural nodality in special degree and denying it elsewhere, the physical lineaments of the earth's surface are highly significant for the growth and density of towns. In one situation, urban growth seems natural, in another well nigh inconceivable.⁴³

Natural nodality, or accessibility, is a major cause of the locations of cities. It is a large part of what is meant by "situation." Cities are often located where junctions of natural routes establish places that are easy to get to. New York, Basel, Copenhagen, and Singapore are cities whose locations are determined by their accessibility on waterways. The locations of Verona, Calgary, Hamadan, and Xi'an are clearly determined by passes. Accessibility is at least one important cause of the location of cities and other settlements. If natural routes exist in regions of slight relief, these too might determine what places are accessible and so cause the locations of cities.

⁴² Brunhes, p. 169.

⁴³ Smailes, p. 56.

The cause of location of cities is a question intimately connected to the question of whether roads follow natural routes. If cities were independent of roads, then the correlation between roads and cities would prove that cities determined roads. In that case, terrain could have only secondary influence. If, on the other hand, roads are independent of cities, the correlation between them can only mean that cities rise on roads.

Site Theory

If there are other conditions, independent of accessibility, that determine the location of cities, these could indirectly determine roads. The principal alternative to the theory of situation is the theory of site, the proposition that local characteristics of the land on which the city is built determine its location.⁴⁴ The theory of site is sometimes given equal weight with the theory of situation, but it cannot provide a consistent and general law.

There are undoubtedly cases in which site factors determine where cities will be. Lowell, Massachusetts and Bolton, Lancashire are located where waterpower is available for their mills. Brighton and Atlantic City are located on good beaches and Damascus is in "the fertile Oasis of Sham."⁴⁵ But there are many aspects to site. Each site factor operates in some cases, but not in others. Frequently they contradict each other. Hill-top sites may be good for defense, but not for water supply. Atlantic City has its beach, but not much solid ground for building. Site does not provide a single, general explanation of the location of cities.

Neither can site be a consistent law. Cities located for site could be at the junctions of roads where roads come to them, but the correlation between cities and road junctions is maintained even where there can be no doubt that the pass points the road. It is no more likely that Innsbruck and Verona determine the location of the road through the Brenner Pass than that Basel determines the course of the Rhine.

Though site factors may influence and even determine the locations of some cities, site is a grab bag of diverse and often contradictory forces that cannot provide a consistent, general law. If there is a general cause of the locations of settlements, these forces can only be exceptions or modifications to it. Though site factors are independent, they do not make cities independent of roads.

⁴⁴ Brian Goodall, *Penguin Dictionary of Human Geography*, (Harmondsworth: Penguin, 1987).

⁴⁵ Vaughan Cornish, *The Great Capitals: an Historical Geography*, (London: Methuen, 1923), p. 57.

Situation Theory

Situation, on the other hand, offers a coherent, general law of cities that explains not only their positions, but their distribution and their very existence. The theory that cities rise on the junctions of roads is completely compatible with Christaller's theory of central places.⁴⁶ Christaller describes the geometric relationships among cities of different sizes. Cities of equal size tend to be equal distances apart with smaller cities evenly spaced between them. According to Christaller, the mechanism that underlies this pattern is the cost of travel. When central places are evenly distributed, the largest possible supporting population is within reach of each town with the lowest possible travel cost.

Christaller's theory explains only relative location, "the number, sizes and distribution of towns."⁴⁷ The same geometric pattern can be established in any position. Some other force is necessary to fix the positions of cities on the land.

The same dynamic of travel cost that Christaller uses to explain the relative positions of cities would also cause cities to be located at the junctions of natural routes. These points can be reached with a lower travel cost and bring a larger population within travel range. The same interest in reducing travel cost between a distributed population and the concentrated functions of a central place underlies the theories of centrality and situation both.

Not only the locations, but also the existence of cities can be explained in terms of travel cost. Central place functions are highly interdependent. Concentrating them in one small area conserves travel cost between one function and another. In addition, since all central place functions share a similar interest in accessibility, they tend to congregate in the same accessible places.

Situation amounts to a coherent general theory of cities in which the locations of cities, the pattern of central places and even the existence of concentrated settlements all spring from the same cause, accessibility. This may be a general cause of the location of settlements. The weakness of this theory is that it yields no explanation for the location of settlements in regions where there are no natural routes. If natural routes can be found in

⁴⁶ Walter Christaller, *Central Places in Southern Germany*, Trans. Carlisle W. Baskin (Englewood Cliffs: Prentice Hall, 1966).

⁴⁷ Christaller, p. 15.

regions of slight relief, situation becomes a general theory that can probably explain the locations of most settlements.

There appears to be no third alternative to the theories of site and situation. J. E. Bird,⁴⁸ offers a list, purportedly exhaustive, of theories of cities. Most of these describe versions of central place functions and fit neatly into the general theory of situation. Others are variations on site theory. Those that remain contain no reference to location and can provide no rational explanation of why a city should be located in one place rather than another.

If a principal cause of the location of cities is their location on roads and other routeways, then cities are not independent of roads and cannot be a cause of road location.

Other Sources of Traffic

There are other places besides cities that might be the destinations of roads. For instance, a road climbs to the top of Pike's Peak. Its goal is not a settlement. Rather the panoramic view is the benefit travellers seek. The benefits that come to people who travel are diverse, involving all the differences there are among places and among people. To conceive of the benefit created by a road as a sum of all these separate benefits is unlikely to lead to a general law. The purposes for which travellers make trips are also beside the point. They are not the same as the purpose of a road. A traveller can reach a goal without using a road. Using a road does not guarantee that he will gain a benefit from the trip. The only benefit that comes directly from a road is that travel takes less work.

When the benefit created by a road is defined as the reduction in the cost of making trips, there are two ways that benefit can be increased. One is to reduce the cost of making trips, but travel cost we have described already. The second is to increase the volume of traffic on the road. The benefit of a road is most usefully measured by its **traffic**, the number of travellers who use it.

Travellers use roads that go where they themselves are going. If there are some places where more travellers go, roads will serve more traffic by going to those places. The

⁴⁸ J. E. Bird, *Centrality and Cities*, (London: Routledge Direct Editions, Routledge Kegan and Paul, 1977).

differences between places that make some the origins and destinations of more trips can be generalized as **traffic potential**. Traffic potential could act on the location of roads.

Most explanations for the locations of roads offered by the geographic literature fit this generalization. Settlements are part of it, but it also includes population,⁴⁹ potential agricultural production,⁵⁰ mineral exploitation,⁵¹ and numerous other conditions. The theory of highway planning is also largely concerned with variations in traffic potential.⁵²

That there is a connection between traffic potential and roads cannot be denied. Taafe, Morrill and Gould have demonstrated a strong statistical correlation between the density of population and the density of roads in Ghana and Nigeria.⁵³ Everywhere, it can be seen that where there are more people there are more roads. But the density of roads is not the same as their location. Roads could have any locations and still maintain the same density. Conversely, the locations of natural routes can be identified without defining how many of these routes are actually occupied by roads. The density of roads is a separate question from the location of roads, and one with which this paper is not concerned.

To define the location of a road is to make a distinction between places at the scale of the width of the road. Rural population and agricultural production are regional characteristics that rarely show sharp edges. Virtually uniform within any small region, the variations in traffic potential that determine the density of roads do not determine their position or alignment.

Other variations in traffic potential are located at specific points. Mines and mill sites, for instance, are small enough to define, in part, the position of a road. Certainly the railroad pattern of West Virginia cannot be explained without reference to the mines, but the pattern formed by mineways is distinct from the pattern of through routes. A mineway is one-ended. It starts from the tittle-head and goes to a network connection. The route it connects to is not itself diverted to the mine. Wherever a road can be found to reach a clear destination, it is normally just such a one-ended road. These roads are not rare, but

⁴⁹ Stanley, p. 419.

⁵⁰ Taafe, Morrill and Gould, p. 388.

⁵¹ Taafe, Morrill and Gould, p. 388.

⁵² Michael J. Bruton, *Introduction to Transportation Planning*, (London: Hutchinson, 1985), p. 15. "The demand for transport is affected by . . . the location of the home, workplace . . . and other activities."

⁵³ Taafe, Morrill and Gould, *op. cit.*

they are almost invariably short. They are not the roads that form the structure of the transportation network. Later we will be able to show that a one-ended road does not usually indicate any specific attraction at its terminal, but for now we will dismiss all such roads as a distinct phenomenon. We will call them **lanes**.

The common idea that roads exist to go to specific destinations describes only the exceptions and not the general case. Since cities are not independent of roads and regional densities cannot define road alignment, there is no variation in traffic potential that can provide a consistent, general and independent cause of the location of roads. The goal of a road is not a destination.

The Area Near a Road

Though it may make no difference which places a road connects, it still makes a difference how many places it connects. The conclusion that variations in traffic potential can have no effect on the location of roads seems to suggest that the benefit side of the economy has no relevance to the question of road location, but this cannot be true. If travellers sought only to minimize cost, they would stay home. An economic equation must have two sides.

As we have said, the benefit created by a road can be defined as traffic, the number of trips that use it.⁵⁴ A **trip** is the action of a traveller in moving from an origin to a destination. The traveller chooses a path. When part of that path coincides with a road, the trip uses the road.

It can be assumed that the normal traveller chooses the **least cost path**. If the origin and destination of a trip are conveniently near the road, the least cost path is to go to the road. But, if the road is far away, it is less work to take a more direct **cross-country** path, and not use the road at all.

Thus the traveller's choice depends in part on the location of the road, and the location of the road depends on how many travellers use it. The circular problem gets complicated, but to find where a road will create the greatest benefit, there is no need to identify which particular trips, or even how many trips will use it. All that is necessary is to describe how changes in the location of a road change the number of travellers for whom the road is part of the least cost path.

⁵⁴ The consequences of this simplistic definition are examined on page 28

A trip as we have just defined it is a single event occurring once. It can be described by a pair of points, but there may be more than one, or no traveller moving between one combination of origin and destination. For any pair of points, there is a frequency with which travellers move between them. This frequency is the traffic potential of the places. We have concluded that no variation in traffic potential can be a cause of the location of roads. Therefore, for the purpose of describing the goal of a road, such variations are irrelevant and traffic potential must be assumed to be constant. Every pair of points is then equally likely to have a traveller moving between them.⁵⁵

The assumption of a uniform traffic potential allows a simple definition of traffic as a geometric area. If the frequency of trips is assumed to equal one, then a trip is the same as a pair of points, and any two points imply a traveller using the least cost path between them.⁵⁶ This makes the traffic of a road proportional to the number of pairs of points between which the road is part of the least cost path. A number of points can be described by an area.

Algebraically, traffic is trips per road. Trips per road equals trips per area times area per road. Trips per area is traffic potential. Since traffic potential is constant, traffic equals area per road. A road that serves a larger area will carry more travellers.

A road serves the area that is conveniently near it. If we assume “conveniently near” can be defined by a simple distance,⁵⁷ the area “near” the road is roughly a rectangle, as long as the road with a width that is twice the distance a traveller will go to get to the road. The limit of distance within which a point is near the road, and beyond which it is not, we will call the **access range** of the road. When there is one trip for every pair of points, the number of trips that use a road is equal to the number of origins within access range of the road multiplied by the number of destinations. The volume of traffic on the road is then proportional to the square of the area of the rectangle.

As long as access range remains constant, rotation or translation of the road does not change traffic volume. Moving a rectangle around on a surface does not change its area. But if the road extends farther, the volume of traffic rises as the square of its extension.

⁵⁵ This assumption is examined on page 49

⁵⁶ That traffic not be infinite, points must be conceived as small finite areas.

⁵⁷ This assumption is examined on page 27

The farther a road extends, the larger is the number of trips for which the road is part of the least cost path.

Extension is not the same as length. If the road is circuitous, it gets longer, but the area near the road does not rise. If the road bends only slightly, the loss of area inside the bend is balanced by a gain on the outside.⁵⁸ The area near a direct road rises as the road gets longer. To create the greatest possible benefit, a road should never end.

Variations in Access Range

We arrive at the conclusion that roads should be endless by assuming that access range is constant. In fact, the distance that a traveller will go to get to a road varies with individual trips, but examination of these variations will not change the conclusion.

The trip made by a traveller who uses a road can be divided in three parts. The first part is an access path from the trip origin to the point of joining the road. The second part is where the traveller is on the road. The third part is another access path from the road to the destination. The cost of the trip is the sum of the costs of these three parts. The traveller following the least cost path will use the road when this sum is less than the cost by any cross-country path.

The cost by cross-country path sets an upper limit to access range. There is no use in going to the road if it is easier to cut through the woods. The total cost on the two access paths may not exceed the difference between the cost on the road and the cost on the cross-country path. This difference defines access range in terms of cost and can be converted into distance by dividing it by the cost per mile of the access trip. The result defines the maximum combined length of the access paths at the origin and destination.

This definition leaves access range highly variable, but one relationship is constant. Access range, and therefore traffic, will rise whenever the cost of travel on the road falls. The easier it is to travel on the road, the farther travellers will go to get to it. Access range also rises when the cost by the cross-country path rises, but both these effects simply repeat the cost concern that overland travel on the line of the road should be easy relative to that on other land nearby.

⁵⁸ This balance remains exact if the bend is circular and the radius of curvature exceeds the access range. It remains nearly true whenever the angle of bend is wide.

Traffic rises when the road extends farther and rises when travel cost falls, but there is no trade-off between these two factors. Extension is far more powerful. For any trip, there is an upper limit to access range. No matter how the cost of travel is reduced, no trip will go to the road if the road is farther away than the destination. There is no upper limit to the length of a road. When a potentially endless road is cut short, the long dimension of the area near it (extension) is reduced far more than the short dimension (access range) can possibly be increased. To create the greatest benefit, a road must, as a first condition, extend indefinitely. Secondly, it should follow the route on which overland travel cost is least, but this second condition simply reiterates the cost side of the economic equation.

Access range may also vary with the orientation of the trip – a traveller is more likely to use a road that goes in the right direction – but this variation is insignificant. If the origin of a trip⁵⁹ is close to the road and the orientations of trip and road are similar, the destination is necessarily close to the road. It is necessarily far away if the orientations of the trip and the road diverge. When every pair of points is a trip, trips of all orientations are so distributed that when the road moves away from some trips, it moves closer to others at the same rate.

The Effect of Distance Decay

The assumption of a uniform traffic potential ignores the fact that short trips are far more frequent than long ones. When two points are far apart, there is less likely to be a traveller between them by a factor of the square of the length of the trip. This is the phenomenon of distance decay.⁶⁰ It is balanced by two effects that make long trips contribute more to the volume of traffic on a road.

The longer a trip is, the more it gains from the easier travel on the road, a saving that can make up for a longer access distance. This means that long trips can go to the road from farther away and are therefore more likely to use the road than short trips are. Thus, long trips contribute more to traffic approximately by a factor of trip length.

Long trips also account for more traffic because they use the road for longer distances. If we use a more sophisticated measure of traffic than the mere number of trips and

⁵⁹ With a length that is not negligible

⁶⁰ Goodall, p. 134.

consider trip miles per road mile, long trips produce more traffic again by a factor of trip length. Though short trips are more likely to occur by two factors of trip length (trip length squared), there are two factors of trip length working to make long trips contribute more to traffic. Together these two effects approximately cancel out distance decay. Equating trips with pairs of points does allow for short trips being more frequent.

Even though the width of the rectangular area near the road is not a simple distance it remains a fair representation of the traffic that uses a road. The normal variations in access range and trip frequency do not affect the conclusion that a road should be endless.

Other Benefit Factors

Another major simplification is to have described traffic as if a road existed in isolation. In fact roads form a connected network, and their locations may be significantly affected by their interaction with other roads. This interaction takes two forms. They compete and they connect. A road will serve a smaller area if it is close to a parallel road. A road will reach a larger area if it enters a junction and forms continuing routes with more than one connecting road. The interaction of roads is a subject we will put aside here, to reopen in Chapter Four where we will find its effects are significant but secondary.

Several other conditions that are not dependent on traffic have been proposed as contributing causes of the location of roads. They include "hegemony,"⁶¹ "unification,"⁶² an "instrument for economic and social progress,"⁶³ and "the desire to connect an administrative center on the sea coast with an interior area for political and military control."⁶⁴ These are all examples of how serving an area of land, connecting it to other places, and making it accessible, can be a benefit in itself. Though not dependent on traffic, this benefit is described by the area near the road and the same conclusions follow.

The theory that roads follow natural routes holds that terrain acts to determine roads through the cost of travel. If any spatial variation were to act on roads in the form of an

⁶¹ Kolars and Malin, p. 124.

⁶² W. R. Stanley, "Evaluating Construction Priorities of Farm-to-Market Roads in Developing Countries: A Case Study," *Journal of Developing Areas* V, no. 3 (1971), pp. 371-400, in Eliot Hurst p. 423.

⁶³ W. R. Stanley, p. 417.

⁶⁴ Taafe, Morrill and Gould, p. 388.

economic benefit, it would contradict the theory. When there is no such factor operating, the benefit created by a road is defined by the area near the road. This rises as the road gets longer.

The theoretical conclusion that a road should be endless agrees with observable facts. Lanes must be set aside, but the through routes that form the structure of the transportation system do not end at junctions or at ports or ever reach destinations of their own. The endless extension of an isolated route defines the benefit side of the economic equation. Marlon Brando, as "The Wild One" speaks for the road when he says, "You don't go anywhere, you just go."⁶⁵

The Hypothesis of Ridge Roads

The conclusion that roads must be endless completes the hypothesis of ridge roads. This is the special case, applying in humid regions of slight relief, of the general hypothesis that roads follow natural routes. It describes early roads that are laid out in largely unoccupied regions and on which the rolling surface is the natural surface of the land.

These roads follow economic routes, routes on which they create the greatest benefit at the least cost. Benefit is described by the area reached by the road. Though the interactions among different roads can have important effects, this area is greatest when the road is endless.

The cost of an early road is described by overland travel cost. This in turn can be described by distance times resistance. Where the shape of the land is dominated by rivers and their valleys, rough ground, soft ground, and steep slopes, the conditions that constitute resistance to early roads, are concentrated along rivers. Other variations in the surface have minor effects but the high ground of the ridge between the rivers provides a firm, smooth, nearly level surface on which travel is relatively easy.

Because distance is a factor in the cost of travel, roads must be direct. There is a trade-off between distance and resistance. The route on which overland travel cost is least is not straight but relatively direct. It does not avoid rivers entirely, but crosses exactly the minimum number it must in order to maintain a nearly constant orientation. It pieces together sections of different watersheds to form a direct ridge route. The physical and

⁶⁵ John Paxton, scriptwriter, "The Wild One" (1954).

economic connection between the ridge and the road should cause early roads to follow ridge routes. That this connection operates, follows from basic physical, mathematical, and economic laws. That it operates with sufficient force to determine the location of roads remains to be seen.

Chapter 3

Constructing a Model of Ridge Roads

To find out whether early roads follow ridge routes consistently requires a predictive model that finds ridge routes. To construct a formal proof of the proposition “Early road if and only if natural route” the hypothesis that has been stated as a physical and economic connection must be expressed as a logical rule. When the minor factors are left out, the hypothesis of ridge roads can be simplified to the following four statements:

- 1) Roads follow routes on which they create the most benefit at the least cost.
- 2) To create the most benefit, a road must be endless.
- 3) Cost equals distance times resistance.
- 4) Resistance varies only to be higher at rivers.

From these four statements and a map of rivers, we will derive a map of ridge routes.

Cartographic Logic

Logical arguments can be constructed in the map. The way a map represents phenomena by areas within boundaries is a basic form of logic. Köppen’s map of climate regions and Thünen’s map of land use areas are logically derived. Their techniques can describe a working connection between conditions and effects where both are spatial patterns. These methods can be adapted to derive a map of natural routes.

Sets and Functions

Köppen's method can be described by set theory.⁶⁶ This branch of mathematics provides precise terms to define the logic of the map. A set is a collection of elements defined by a rule. It is a logical category, the things of which some statement is true. A bounded or colored area on the map is a set of points on paper that corresponds to a set of places with some common characteristic on the land. In Köppen's system, the statement, "this place averages more than 64 degrees in the coldest month" defines the category of hot places and the isotherm defines which points are contained in that set.

The rule that defines a set is a function. It assigns elements of one set to elements of another.⁶⁷ Temperature is one example of a function. It assigns points on the thermometer scale (elements of the set of temperatures) to places with thermometers (elements of the set of places). The domain of this function is all the places at which temperature can be measured. Its range is the centigrade or Fahrenheit scale on the thermometer.

Temperature is a continuous function which means simply that if it is 100 degrees in Stanleyville, it is near 100 degrees anywhere near Stanleyville. If a function is continuous, the set of points at which it has a common value form a closed boundary with all points on one side having values higher than those on the other side. An isoline, or contour, can be constructed for any continuous function, including elevation, rainfall and barometric pressure. A contour of rainfall (isohyet) of 2.4 inches in the driest month defines Köppen's wet region.

Sets defined by boundaries are Venn diagrams. Venn diagrams can also define the logical connectives "and," "or" and "not" that are relations between categories, and from which all valid inferences are constructed (figure 1). The connective "and" is the intersection of two sets, the area where the two boundaries overlap. Köppen's Af climate region is the

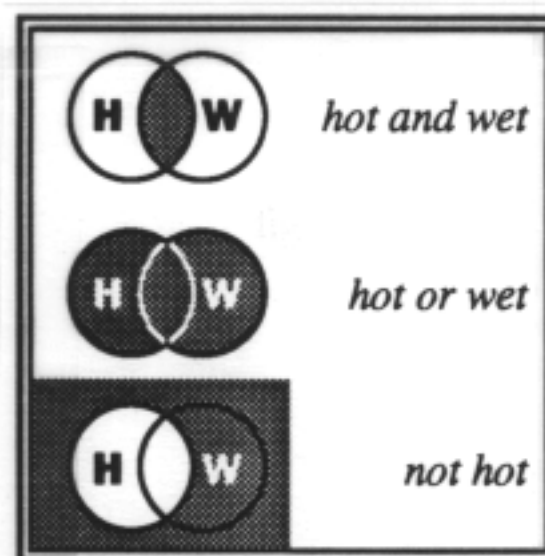


figure 1

⁶⁶ Murray Eisenberg, *Axiomatic Theory of Sets and Classes*, (New York: Holt Rhinehart and Winston, 1971). (We modify his terminology in some cases, specifically in using "slice" for the more formal "partition.")

⁶⁷ Mathematical equations are functions whose domain and range are sets of numbers.

intersection of the hot region and the wet region. Negation is the area outside the boundary. Logical negation is symmetrical, so it can be an arbitrary choice which side of the boundary to call "inside" and which "outside." "Not dry" is actually the first step in Köppen's system. The union of two sets, the area inside either one boundary or both, defines the connective "or." Using the logic of the Venn diagram, arguments of considerable complexity can be constructed with boundaries in a map.

A map is analogous to a second logician's device, the truth table. This describes all possible cases as boxes of a grid and defines for each box whether the statement is true in that case. If the statement is true the box is filled with a T; if false, with an F. If the truth table of P is identical to the truth table of Q, "P if and only if Q" is true. A map is a truth table that substitutes colors for letters. Each point of the map is a case, an infinitesimal box with the color of the map indicating the truth value of some statement. Superimposing two maps compares two truth tables and is a formal logical proof. The correspondence between Köppen's Af region and the characteristic vegetation of the tropical rain forest is, in effect, a proof that the jungle is hot and wet. Though the correspondence is not perfect, to demand the exact match that abstract philosophy expects would be to reject the use of logic in practical affairs.

Roads Described by Sets and Functions

Roads can be described by bounded areas. The roads on a map are usually areas wider than the true scale width of the road. Constructing a map of routes means constructing boundaries to the route areas that contain route curves.

The historic process of road development, the act of road location, is a function. Its domain is the surface of the land and its range is the two sets of points, those in roads and those not in roads. Putting roads on the land puts points on the land into the set of points in roads. A model that describes this process is also a function. We will call it the **natural route function**. It should assign the same points to the same sets as does the historic process.

The assumption that roads follow economic routes means that the natural route function can be conceived as the economic equation of costs and benefits. The natural route is a solution to this equation. The equation has some value at any point on the

surface. Its value is low in swamps and on mountaintops. Natural routes are where the value of the natural route function is highest.

The natural route function is continuous. Small changes of route make only small changes in the cost of travel. If one point is near another point, its value as a road is near that of the other. Only where there are sharp discontinuities of terrain, as where one false step goes off a cliff, is this not true. Because it is

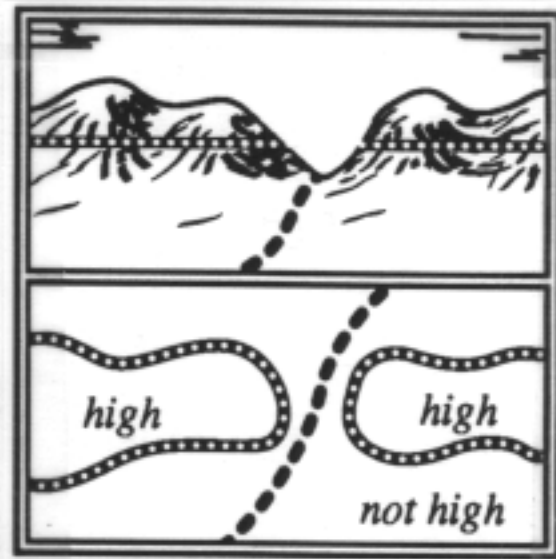


figure 2

continuous, the natural route function can be described by contours. The local maximum of a continuous function is always on the high side of any contour that describes that function. Since a natural route is a local maximum of the natural route function, contours of this function will find natural routes.

In mountain areas, contours of elevation are contours of the natural route function. The value of a route is highest where the land is lowest. The area outside the contour contains the natural route (figure 2). Outside mountains, elevation no longer determines roads and the equation of costs and benefits is harder to solve. There is, however, no need to define its absolute value. It is only necessary to define the difference between one route and others near.

As Wellington says, speaking of railroads:

It is the duty of the engineer to neglect [whether the profits will be large or small] absolutely in laying out his work, considering only the effect of his decisions upon these three items:

1. THE DIFFERENCE in gross receipts which will or may result from choosing one or another line.
2. THE DIFFERENCE in operating expenses which will or may result from choosing one or another line, one or another gradient, one or another limit of curvature, etc.
3. THE DIFFERENCE in annual interest charge which will or may result from the differences in cost of construction caused by differences in the above details.⁶⁸

⁶⁸ Wellington, p. 17-18, emphasis in the original.

The case is similar when the choice is made less systematically by Indians or frontiersmen. The order, not the magnitude, of costs and benefits, defines the natural route. Contours of the natural route function need describe only the difference between routes.

The benefit side of the equation is defined by the condition that a road must be endless. On a finite map, an endless route must cross from edge to edge. The converse is not necessarily true, a line across the map need go no farther. Crossing the map is a necessary but not a sufficient condition, but, if the region is large enough, it is a virtually sufficient definition of how a road will create the greatest benefit. What Wellington called “gross receipts”⁶⁹ can thus be defined with an absolute condition. Among the routes that meet it, the only significant differences are their costs. Since all the costs of an early road follow the cost of overland travel, contours of the natural route function are contours of relative overland travel cost. They describe the equation cost equals distance times resistance.

Graphic Construction of Contours

Thünen demonstrates how contours can be constructed that describe an economic equation.⁷⁰ He derives a map of agricultural land use areas by graphing an equation and geometrically projecting the graph into a boundary. Though his method may be familiar to the reader, we describe it in detail to establish terms and concepts that we will later use in the case of natural routes. Thünen’s first step is to construct a graph of the function that relates land use to location. The range of this function is net profit on production, which he defines as equal to market price less transport cost.⁷¹



figure 3

This is on the vertical scale (the abscissa) of the graph. The domain of the function, on the horizontal scale (the ordinate) is distance from the market city.

Assuming travel in a straight line on a uniform plane, travel cost, in Thünen’s model, is described by a straight line rising with distance from market. Price minus transport cost is a straight line starting at market price and falling with distance (figure 3). This graph assigns

⁶⁹ Wellington, p. 17.

⁷⁰ Thünen, op. cit.

⁷¹ Production cost of each product is an assumed constant.

to each distance from the city a net profit on one commodity. Because different commodities have different prices, and also different transport costs, their profit curves intersect. Wood has a high price and a high transport cost, so its curve starts high and falls fast. Wheat has a lower price and a lower transport cost, so its curve starts low and falls slowly, and crosses the curve for wood.

The intersection of the profit curves defines a boundary point on the ordinate. Inside this point the profit on wood is more than on wheat. Outside, wheat yields more than wood. The boundary point is a solution to the economic equation that describes the choice between commodities. Since distance from market is a location coordinate in the polar system with its origin on the market city, the boundary point is a solution defined in terms of location.

This solution can be re-expressed as map areas. When the ordinate line is rotated around the market center, it generates the plane of the map, and the boundary point generates a circular boundary (figure 4). Inside this circle, wood is produced, and outside, wheat. By graphing the economic function that connects profit to location and projecting the graph to generate a boundary in the plane, Thünen logically derives a map.

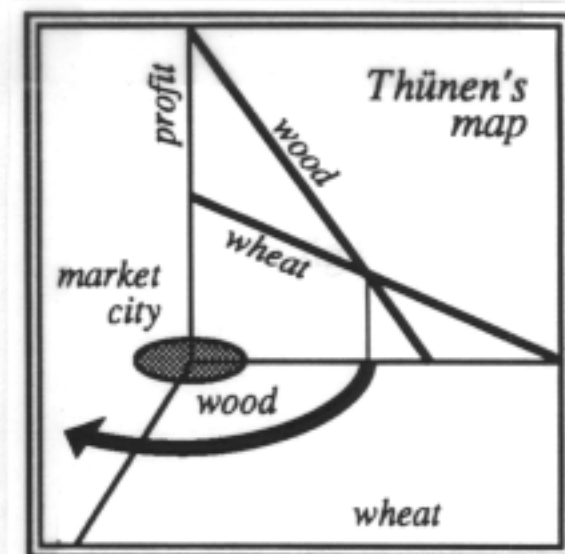


figure 4

A Graph of Travel Cost

Thünen's method can be adapted to derive a map of ridge routes. Travel cost is defined by the equation distance times resistance. This equation can be graphed, and the graph projected into a contour of travel cost.

In part our problem is simpler than Thünen's. We are interested in only one "commodity," the normal traveller, but two of Thünen's simplifications cannot be used. Since we want to explain where roads go, we cannot assume a given destination, and, since we must explain where and why roads bend, we cannot assume travel in a straight line.

Travel cost is a characteristic of paths. The work required to move a load can only be defined when distance is known, and this requires that the endpoints and alignment of a

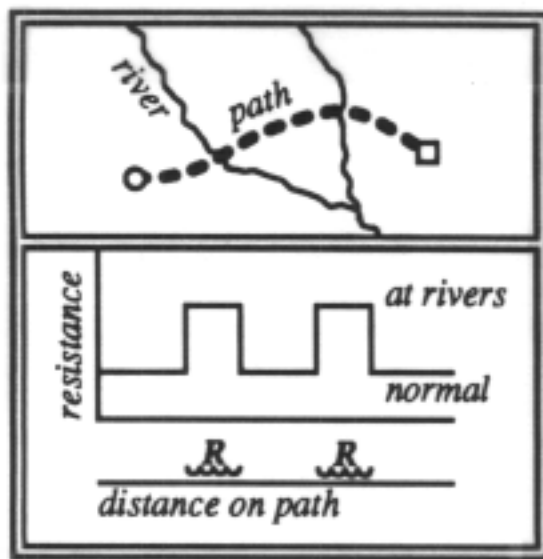


figure 5

path must be given. The resistance of the surface passed over must also be known and this too requires that alignment be defined.

The assumption that resistance varies only to be higher at rivers, means that this variable can be read from a map of the drainage system. This assumption also means that all areas not at rivers have a uniform resistance, not low, but lower than near the rivers. We will call it **normal resistance**. This surface is in effect,

a Euclidean plane.⁷² On this plane, distance can be directly measured or, as will prove more useful, it can be described by geometry and trigonometry.

A graph of the travel cost equation can easily be constructed for a single path. If we assume the position and alignment of a continuous curve with endpoints, the ordinate of the graph can be defined as the scale of distance along this path. Each point of the ordinate corresponds to one point on the map. If the path is found on the map of rivers it can be seen which points, if any, are at rivers. Since resistance can be equated with rivers, this defines a graph of resistance (figure 5). The graph curve is horizontal through all points not at rivers, where resistance is a normal constant. It shows peaks at river crossings.

The graph of resistance can be multiplied by distance from the origin. This gives a graph with location on the ordinate and cumulative overland travel cost on the abscissa (figure 6). The curve rises at a constant slope through all the points with normal resistance and at a steeper slope wherever the path crosses a river. The altitude of the curve at the destination defines the total cost of the trip on the assumed path. The vertical scale of this graph is not defined, but the shape of the graph is similar whatever the vertical scale, provided that normal resistance is

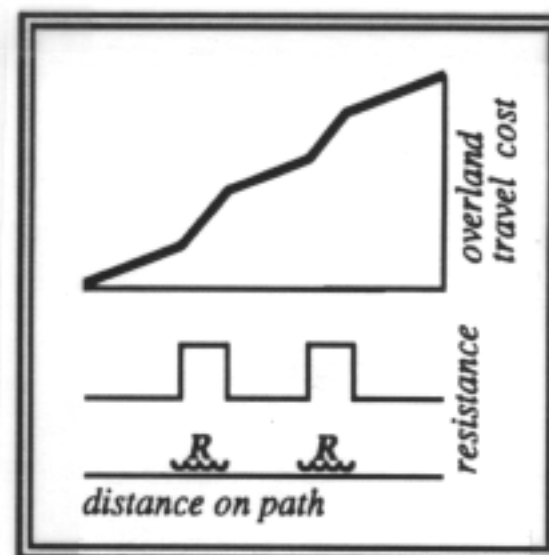


figure 6

⁷² "Non-Euclidean time distances . . . to make allowance for . . . variable terrain conditions" (C. C. Kissling, "Linkage Importance in a Regional Highway Network," *Canadian Geographer* XVIII, 2 [1962] pp. 113-129, in Elliot Hurst, p. 93.) are simply the effects of variations in resistance that are here accounted for by the map of rivers.

constant and resistance at rivers is higher. This graph of overland travel cost translates terrain conditions as read from a map of rivers into a definition of overland travel cost. It begins the process of reading relative travel cost from the map.

Routes as Sets of Paths

Conceiving of a route as a set of paths defines the problem we face as selecting from the set of all paths, those that are in natural routes. We will define a **natural path** to be a path that coincides with a segment of a natural route. Finding natural routes is defining the set of natural paths. These are contained in the set of all paths, which is all continuous curves between all pairs of endpoints (figure 7).

As a first step to finding natural paths, we will assume that two points on a natural route are given. This is, of course, to assume that which is ultimately to be derived: what points are in natural routes. This assumption can later be eliminated, starting on page 49. But first we will define which of the many paths between two given points could be a segment of a natural route.



figure 7

The alignment, or map shape, of a route segment cannot affect the extension of the road beyond its ends. The benefit side of the economic equation has no bearing on local alignment. Only the cost of the road, defined here by overland travel cost, can influence the choice of route alignment. A segment of a natural route must be the path with the least overland travel cost. We will define the **least cost overland path** to be the path that a traveller will follow if he moves on the natural surface of the land from the given origin to the given destination and does it with the least possible work. Though there are many paths for any trip, there is presumably a unique path on which the cost of overland travel is lower than on any other. A natural path must be the least cost overland path between its own endpoints.

There is a least cost overland for any pair of points. Such a path is a natural path only if its endpoints are in a natural route, but we will be halfway to finding natural paths when we can locate the least cost overland path for any trip

Components of Travel Cost

Only the difference in cost need be defined to identify the one path with the lowest cost. If all paths for one trip can be put in order by relative overland travel cost the least cost overland path will be found. We have constructed a graph of travel cost, but it describes only one path. Using it to compare different paths requires more development. The cost of a trip can be analyzed into three parts, separating the components that are attributable to straight line distance, to detour, and to river crossings.

The cost of a trip depends first of all on the straight line distance between its origin and destination. We will call this **trip length**. Path length may be more than trip length, but it cannot be less. Nor can resistance be lower than the normal resistance of land away from rivers. Therefore, there is, for any trip, an absolute minimum cost equal to trip length times normal resistance. This we will call the **base cost** of the trip.

Paths that bend and paths that cross rivers have total costs that exceed the base cost of the trip. The difference between the total cost and the base cost we will call the **excess cost** of the path. Any two paths for one trip have the same base cost, so if they are put in order by excess cost they are in order by total cost.

Part of the excess cost of a path is attributable to detour. If a path is not straight, it is longer than the trip. If it crosses no rivers, its total cost is equal to path length times normal resistance. The difference between

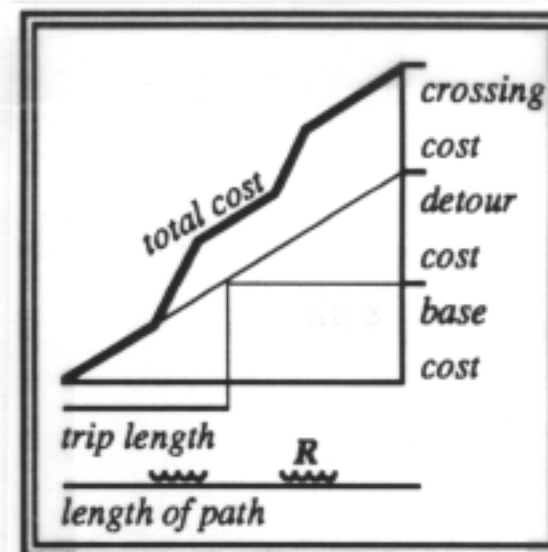


figure 8

this cost (the minimum for a path of given length) and the base cost of a trip we will call the **detour cost** of the path.

The rest of excess cost is attributable to river crossings. The difference between the total cost of a path that crosses rivers and the minimum cost for a path of the same length, we will call the **crossing cost** of the path.

Total cost, or overland travel cost, is equal to base cost plus excess cost. Excess cost is equal to detour cost plus crossing cost (figure 8).

Crossing cost is equivalent to a count of river crossings. If every river were the same, the cost of all the river crossings in a path would be equal to the number of rivers times a constant. The number of crossings would describe crossing cost. This would still be true

if the cost of crossing rivers were nearly constant. The order defined by the number of crossings could differ from the order defined by the actual crossing cost only if some river were far more difficult to cross than were all other rivers. If we assume that large differences between rivers are not the normal pattern of humid regions of slight relief, paths ordered by a count of crossings are ordered by crossing cost. A count of river crossings can easily be produced for any path found on the map of rivers.

This analysis of travel cost into detour cost, crossing cost and base cost makes it possible to compare travel cost for different paths on the basis of conditions read from the map, but it remains severely limited. So far it can make only the trivial observations that if two paths are the same length, the one crossing fewer rivers is easier, and that if two paths cross the same number of rivers, the shorter path is easier. Identifying the least cost overland path means putting all paths in order by cost. To do this, it is necessary to distinguish between one path that is longer and another that crosses more rivers. This means solving the balance between detour cost and crossing cost.

Solving the Distance Resistance Trade-off

The balance between detour cost and crossing cost is a restatement of the trade-off between distance and resistance. A traveller with a river ahead of him faces a choice to cross it or to go around. Straight ahead, detour cost is zero, but crossing cost is high. Going around the headwater eliminates the crossing cost, but increases the distance travelled. Choosing the route with the lowest overland travel cost is a long series of choices between crossing rivers and going around headwaters. Each choice may be described by the question, how much extra distance is equivalent to the work of crossing a river? The answer depends on the difference between detour cost and crossing cost.

The graph of overland travel cost, as yet defined (total cost in figure 8), cannot compare detour cost with crossing cost. The slopes that represent the higher and lower levels of resistance have been arbitrarily assumed. Total cost is defined by the altitude of the graph, but this changes (figure 9) with

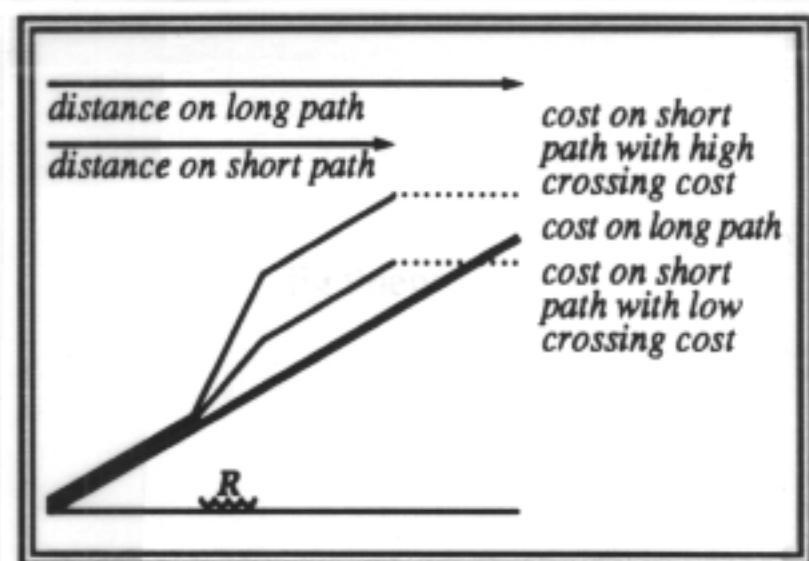


figure 9



figure 10

different assumptions about its slope. If the two levels of resistance are nearly the same, a shorter path crossing rivers will have the lower graph. If the two slopes are widely different, the graph for the shorter path will contain very steep sections at the rivers. The total cost on the short path will then be higher than on the long one. The difference between the two slopes defines the balance between detour cost and crossing cost. To define the correct slope of the graph is to

solve the trade-off between distance and resistance.

To compare the costs of detour and river crossings, they must be expressed in common terms. This might theoretically be accomplished by defining the vertical scale in monetary terms, but we are dealing with costs that were never measured this way. The available evidence is the choices that were made in the act of locating roads. This evidence is in the form of mapped locations. Both crossing cost and detour cost are dependent on location, and the solution is to be expressed as a location. Location is the common term in which the balance of costs must be solved.

Thünen's graphic method solves an economic equation in terms of location, but his graph has a range of different locations on the ordinate; our graph has only one, a single path. To adapt Thünen's method to solving the trade-off between distance and resistance, it is not the vertical scale of the graph that should attract our attention, but the horizontal. If a graph of travel cost can be constructed with a range of different locations on the ordinate, the least cost overland path will be found at the lowest point of the cost curve. The difficulty in constructing such a graph is that the locations are paths. The set of all paths for one trip do not fall easily into a linear order.

Paths of Simple Shape

We have defined the set of all paths as all curves between endpoints, but this definition is unnecessarily broad. Though discontinuous paths and paths that leave the surface are excluded, it still includes calligraphic flourishes that would make ridiculous roads (figure 10). The routes of roads are simpler than this definition allows for. Route

segments can be approximated by simple geometric shapes. A set of paths of simple shape can be arranged on the ordinate of a graph.

A path of simple shape can be built from two sides of an isosceles triangle. We will define an **isosceles path** as a continuous curve formed from two equal straight line segments (figure 11). The base of the triangle is the straight line of the trip. The angle between the path and the straight line, the base angle of the triangle, we will call the **angle of detour**. If the origin and destination of a trip are given, the angle of detour defines the exact position of the entire isosceles path. Because its base angles are equal, an isosceles triangle is completely defined by one side and one angle.⁷³

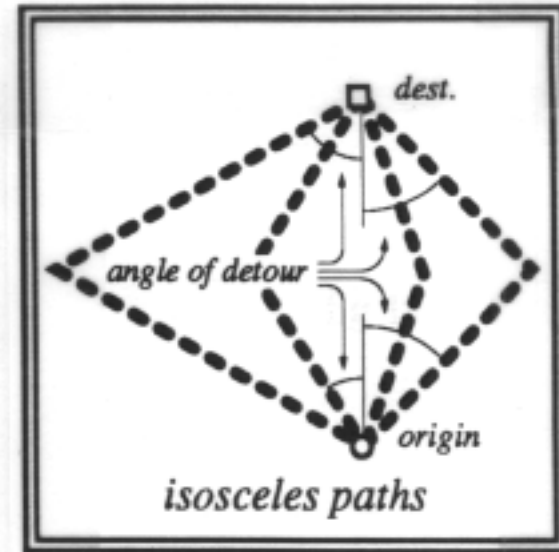


figure 11

Because its base angles are equal, an isosceles triangle is completely defined by one side and one angle.⁷³

Isosceles paths fall naturally into a linear order. A graph can be constructed with a scale of angular measure on the ordinate. Each point on this line corresponds to one path in the set of isosceles paths for a given trip. With the angle of detour on the ordinate, detour cost and crossing cost can be plotted for the set of all isosceles paths for one pair of points.

A Graph of Crossing Cost

The graph of crossing cost is simple. When crossing cost is roughly the same all along the river, it is a single constant for all paths that cross the river. It is zero for all paths that avoid the river (figure 12). If a river lies between the origin and destination of the trip, there is some isosceles path that passes through the headwater of the river. This path has an angle of detour that defines one point on the ordinate of the graph. All paths inside it have a smaller angle of detour and cross the

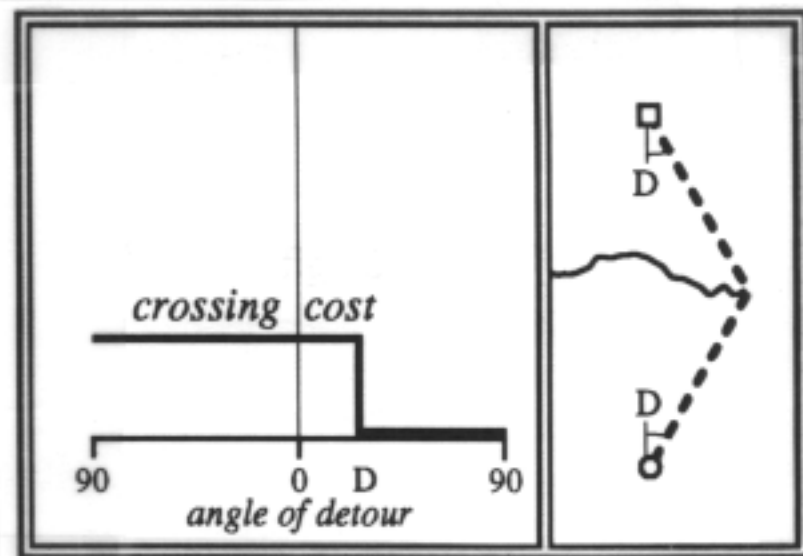


figure 12

⁷³ The triangle has a mirror image with the same base and the same base angle, but this can be distinguished by defining it with a negative angle of detour.

river. All paths outside have a larger angle of detour and avoid the river. The graph of crossing cost is a horizontal constant that drops to zero at the angle of detour of the path through the headwater.

A Graph of Detour Cost

The graph of detour cost is trickier. Detour cost can be described by a ratio. We will define the **detour ratio** as the length of the path divided by the length of the trip. Paths for one trip that are in order by detour ratio are in order by detour cost. This ratio can be defined by a trigonometric function. In an isosceles triangle, the length of one side divided by half the base is the secant (hypotenuse over adjacent) of the base angle. Therefore the detour ratio of an isosceles path is the secant of the angle of detour.

The graph of the secant function is a U-shaped curve (figure 13). The center of the

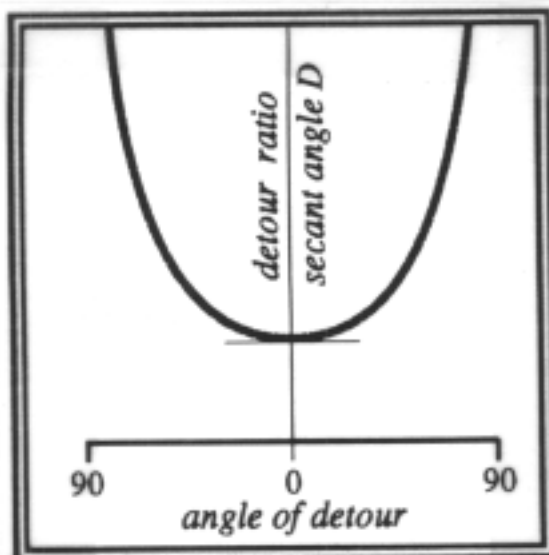


figure 13

graph is where the angle of detour equals zero and the path described is a straight line. There the detour ratio is one: the length of the path equals the length of the trip. At the center, the graph is horizontal. It then rises very slowly through small angles before it accelerates and finally reaches infinity at ninety degrees.

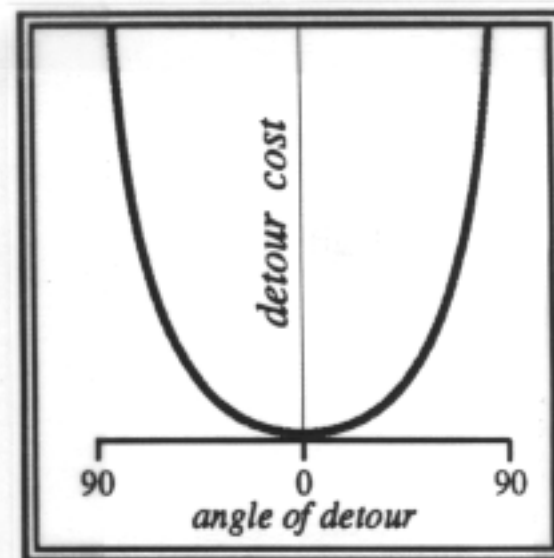


figure 14

The shape of the secant curve provides a mathematical explanation of why a taut string is so easily pulled out of line (see page 18), and why roads are seldom straight. If a path bends slightly, it remains

in the nearly horizontal section of the curve and its length increases by only a trivial amount. At thirty degrees, an isosceles path is only fourteen percent longer than the straight line. Detour cost is significant only on paths that diverge far from the straight line.

The secant curve does not describe detour cost directly. The vertical scale shows the detour ratio, not detour cost. Its value is one on the straight line where detour cost is zero. Subtracting one from the secant moves the center of the graph down to zero (figure 14). This is the effect of assuming trip length to be constant. The graph of secant

minus one is a graph of detour cost, with a vertical scale that incorporates normal resistance and the constant trip length. The units of the vertical scale are undefined.

The Least Cost Overland Path Found

The least cost overland path is to be found at the lowest point of a graph of excess cost. Detour cost and crossing cost can be graphed separately, and their sum is excess cost, but the two graphs cannot be added because their vertical scales are not the same. With a common vertical scale assumed, however, the two graphs can be superimposed, and their interaction then describes the choice between crossing a river and going around it.

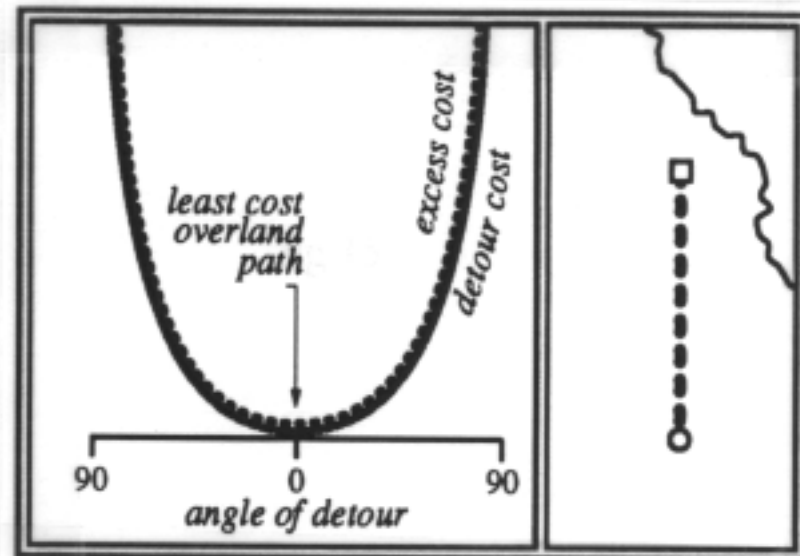


figure 15

If there is no river near, crossing cost stays at zero and excess cost is the same as detour cost. The low point on the graph is at the center of the curve where the angle of detour and detour cost both equal zero (figure 15). The traveller taking the easy way goes straight ahead.

If a river lies between the origin and destination, and the headwater is far away, the graph of excess cost is similar to the graph of detour cost but elevated, with its minimum equal to crossing cost. The low point of the graph and the least cost overland path are still at the center of the curve, on the straight path where excess cost equals crossing cost (figure 16). When the headwater is far away, the least cost overland path crosses the river.

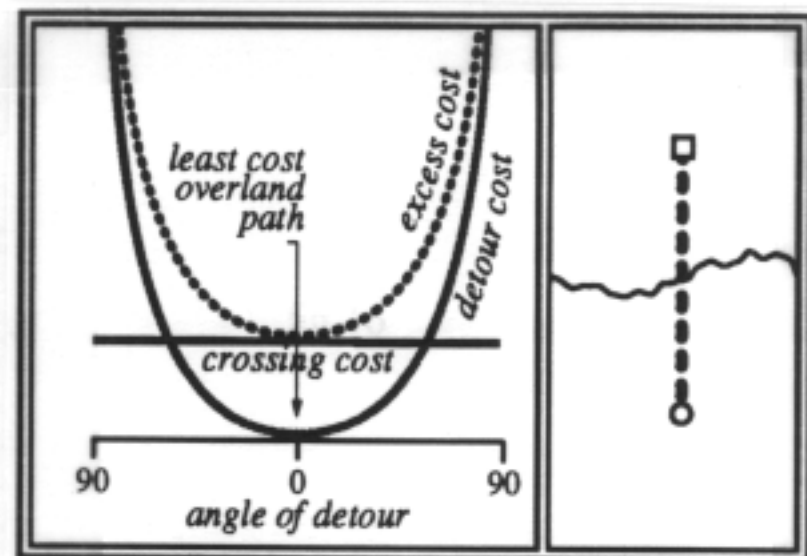


figure 16

If, on the other hand, the headwater of the river is near, the graph of crossing cost drops to zero before detour cost rises across it. Below the headwater, excess cost is the elevated secant curve, while beyond the headwater excess cost follows the detour cost curve (figure 17). The low point in the graph

of excess cost is at the path that bends around the headwater. Here excess cost is equal to detour cost and is less than the cost of crossing the river. When the headwater is near, the least cost overland path bends to avoid the river.

The Limit of Detour

Superimposing the two graphs of detour cost and crossing cost defines how near the headwater must be before it is easier to avoid the river than to cross it. If the horizontal of crossing cost is extended, it intersects the curve of detour cost (figure 18). The point of intersection defines the angle of detour at which detour cost equals crossing cost. We will

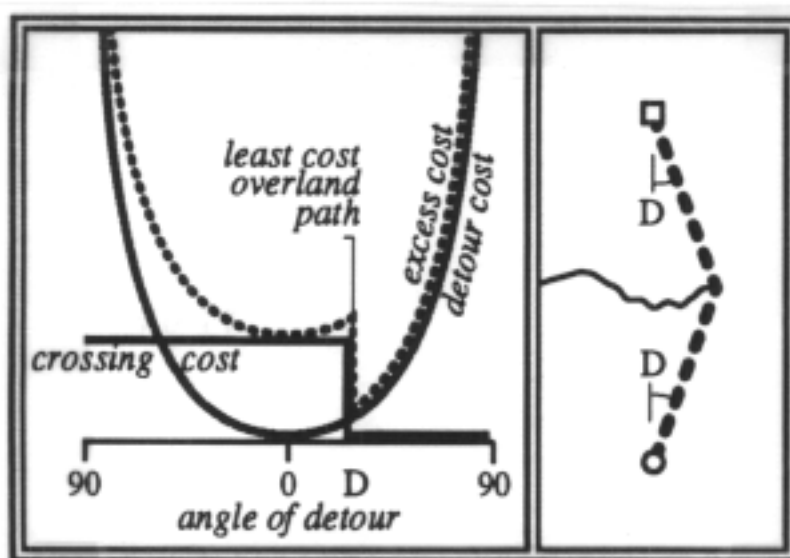


figure 17

call this angular measure the **limit of detour**.

If the angle of detour of an isosceles path is greater than the limit of detour, detour cost is greater than the cost of crossing a river. Such a path has a higher total cost than the straight path crossing the river. It cannot be the path with the lowest travel cost. If an

isosceles path is the least cost overland path for a trip, it must have an angle of detour less than the limit of detour.

The measure of the angle at which detour cost equals crossing cost depends on the arbitrarily assumed vertical scale of the graph. Conversely, the angle defines the scale. The limit of detour can define the ratio between the two levels of resistance, the normal resistance of land away from rivers and the higher resistance of the river and its valley. Though we do not yet know the measure of this angle, the limit of detour defines a solution to the trade-off between distance and resistance in terms of angular measure. The limit of detour can solve the central problem in the location of roads.

A Direct Path Defined

The limit of detour defines "direct." A **direct** isosceles path has an angle of detour less than the limit of detour. Only if it is direct can an isosceles path be the least cost overland path between its endpoints. Therefore, a segment of a natural route, if it approximates an

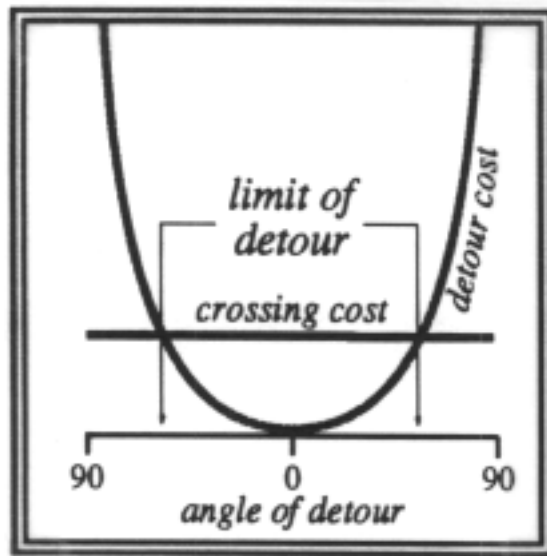


figure 18

isosceles path, must be direct by this definition. The limit of detour defines how much a route can bend and still be the route with the lowest travel cost.

That the limit of detour is constant throughout a region follows from assumptions already made. If all rivers are equally difficult to cross, the horizontal of crossing cost is fixed on the vertical scale, and if travel cost away from rivers is constant, the secant curve is also fixed.

The limit of detour as defined by the graph of excess cost varies with the length of the trip.⁷⁴ The constant crossing cost is equal to proportionately longer detour in a short trip than in a long one. Short trips can take sharper bends, but the bends in a road cannot vary with the length of trips. If there are roads that carry only short trips, these could bend more than through routes, but a single limit of detour should be expected to describe all bends in all main roads under reasonably homogeneous economic and terrain conditions.

There is no way to calculate what angle defines the limit of detour. The evidence of the balance of costs is only to be found in the locations of roads, the choices that were made by those who directly perceived the costs of travel. If historic roads follow natural routes, no road will bend by more than the limit of detour. The angles in roads can be observed and measured. Early roads rarely diverge from a straight line by more than about fifteen degrees.

The exact balance of costs is immaterial. The choice of whether to cross or go around a river is a choice between limited alternatives. If the limit of detour describes even an approximation of the balance, it will be easy to decide which is the low cost choice. Only in the rare instances where the route around the headwater is almost exactly at the limit, will a small change in the limit of detour change the answer.

An angle of thirty degrees is generous enough to be sure that most early roads fall within it, yet not so wide that it includes circuitous routes. Partly because it is easily constructed and convenient to work with, we will define the limit of detour at thirty degrees. A direct path has an angle of detour less than this limit.

⁷⁴ A trip length was assumed in constructing the graph.

The Rhombus of Direct Paths

The limit of detour defines a direct path in terms of angular measure. To find direct routes on the map, this definition must be expressed in terms of map location. With the origin and destination of one trip plotted on the map, isosceles paths at the limit of detour can be constructed between them. If one path bends left and the other right, the result is a rhombus, a figure with four equal sides (figure 19). The angles on the endpoints of the

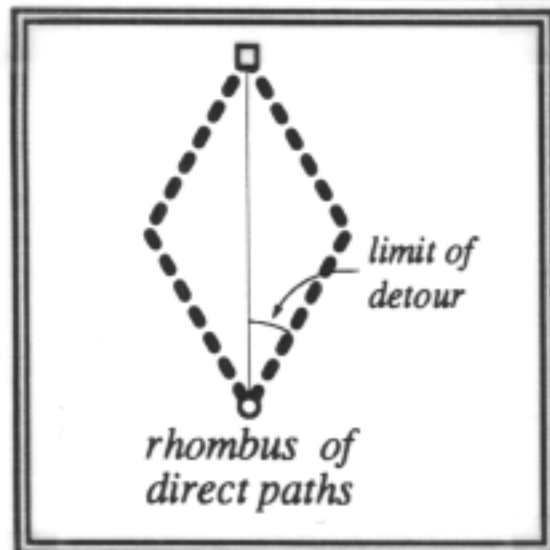


figure 19

trip measure twice the limit of detour, or sixty degrees. We will call this the **rhombus of direct paths**. All direct isosceles paths for the trip between these endpoints are contained within this figure.

The rhombus of direct paths describes only isosceles paths exactly, but it approximately describes all direct paths of any shape. Any path can be described by the detour ratio (path length divided by trip length). If the detour ratio is less than the secant of the limit of detour, the path is no longer than a direct isosceles path and can be called a direct path. All direct paths so defined are contained in an ellipse with its foci on the the origin and destination of the trip. The construction of an ellipse with string illustrates this. If the pins are in the trip endpoints, the length of the taut string defines paths with the maximum detour ratio. The moving pencil describes their extreme range. All shorter paths for this trip correspond to loose positions of the string inside the ellipse.

The rhombus of direct paths is inscribed in this ellipse (figure 20). There is no difference between the rhombus and the ellipse in the critical dimension perpendicular to the trip, and only small differences elsewhere. The solution defined for isosceles paths closely approximates the solution for paths of any shape. Virtually all direct paths for any one trip are contained in the rhombus of direct paths constructed on the endpoints of that trip. This figure locates paths on which the cost of detour is less than the cost of crossing a river. The rhombus constructed with sixty degree angles on the origin and destination of a trip defines the approximate location of the least cost

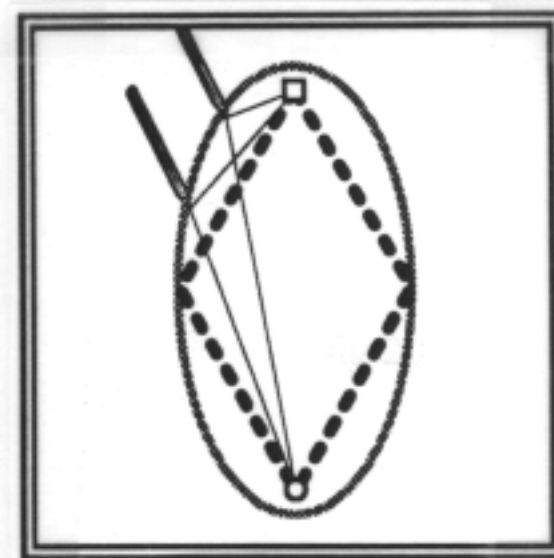


figure 20

overland path for that trip. This is the one path between them that can be a segment of a natural route.

Routes as Sets of Trips

The rhombus of direct paths narrows the set of all paths to a far smaller set that contains all natural paths. We can reasonably say that it defines one path for any pair of points, even though it does not define the exact location of that path. With the assumptions of a uniform traffic potential and normal traveller that we have already made, a trip has no other characteristics than its two endpoints. "Trip" is a synonym for "pair of points." Since any path has a unique pair of endpoints, the set of all paths can be thought of as the set of all trips multiplied⁷⁵ by the set of paths for each trip. Since the latter set has been reduced to one, the least cost overland path, all that remains to selecting natural paths from all paths is to select, from the set of all trips, those that are

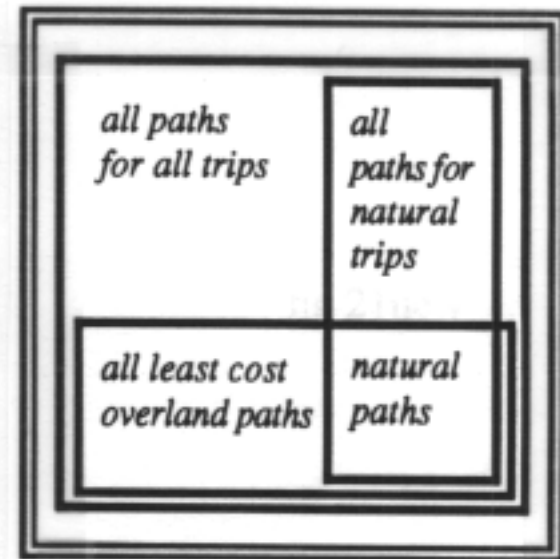


figure 21

endpoints of natural paths (figure 21). We will define a **natural trip** as a pair of points that are the endpoints of a natural path. Every pair of points in a natural route is a natural trip. To find natural routes means to select, from the set of all trips (all pairs of points) the set of natural trips. The temporary assumption of given points that we made on page 39 can now be removed.

Defining the set of natural trips is simply a restatement of the problem of defining the set of points in natural routes, made more complicated by considering points in pairs. It is, however, a useful restatement. Travel cost as an effect of terrain conditions can be defined for paths. Because the rhombus of direct paths locates one path for any trip, cost can be defined for trips, and trips can be compared.

So far we have used the rhombus of direct paths to compare different paths for one trip. To use it to compare travel cost for different trips requires one more simplifying assumption. We will assume that detour cost is zero on any direct path. Since the detour cost curve is nearly horizontal at its center, the cost of detour cannot be significantly

⁷⁵ According to Eisenberg (Op. Cit.), sets cannot be multiplied and the correct term is the "grand union" of sets.

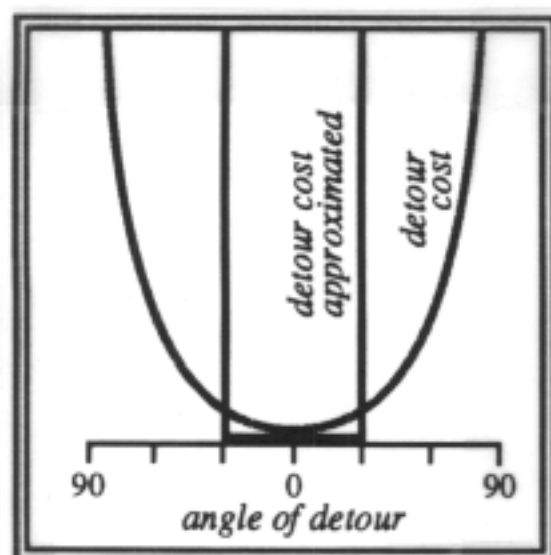


figure 22

reduced by making a direct path straighter. This assumption is the same as to approximate the secant curve with two vertical lines (figure 22). It eliminates relative degrees of directness. Any two direct paths (for trips of equal length) are in order by total travel cost, if they are put in order by crossing cost.⁷⁶

The rhombus measures whether any trip has a direct path that avoids rivers. A rhombus constructed on the map locates the least cost overland path of a trip so it can be seen whether it crosses a river. If a river crosses the rhombus all direct paths for that trip must cross the river. If a river enters but does not cross the rhombus, there is a direct path for that trip that does not cross the river, and the cost of that trip is as low as if no river were near. The rhombus of direct paths defines a set of trips that we will call **easy trips**. An easy trip has some direct path that does not cross a river. The cost of travel on the least cost overland path of an easy trip is as low as the cost on any path for any trip of equal length. The rhombus of direct paths measures relative travel cost from the map of rivers. It can be used to compare different trips and decide which are easy.

The rhombus of direct paths finds easy trips, but a model of roads must find natural trips. Since ridge routes may cross some rivers, they must have segments that cross rivers. Some natural trips are not easy trips. Though we can find easy trips, it is also necessary to examine how trips accumulate to form endless routes on which overland travel cost remains low throughout. Whole routes must be compared, and not just segments of routes.

Truly endless routes are hard to handle, but routes that cross the map are a practical approximation. Such a route is not necessarily endless, but an endless route must cross the map from edge to edge. Crossing the map defines a route that creates the maximum possible benefit. Since a natural route is where a road creates the most benefit at the least cost, it can be defined as the route that crosses the map with the lowest overland travel cost.

⁷⁶ Total cost = base cost + detour cost + crossing cost. Base cost is equal for any two trips of equal length.

A route across the map is actually a path. It has endpoints at the edges of the map, and it has a definite travel cost. But a rhombus of direct paths constructed on its endpoints is too wide to locate a route with any precision. The rhombus can measure only short paths, but long routes can be built up from these.

Since every segment of a ridge route is direct, the route as a whole must be direct. All segments of a **direct route** remain within thirty degrees of the straight line describing the general orientation of the route.⁷⁷

How natural trips, or pairs of endpoints of natural paths, accumulate to form long routes can be described by a route area. When paths accumulate to form a route, their endpoints accumulate in a continuous curve. Boundaries of a route area define an area that contains endpoints of natural paths. We will construct contours of travel cost that have endpoints of easy trips only on one side. From these contours, other contours can be constructed that have endpoints of natural paths only on one side.

The Problem of Dimensionality

Constructing contours of travel cost runs up against the major limitation of the map. A map puts ink on paper to correspond to the characteristics of the Earth's surface. It can put no more than one color of ink at each point. For a condition to be described by a map, it must have no more than one value at each point. This makes it difficult to map travel cost. There is never a single value of travel cost at a point. Every point is the origin of trips to every other point on the map, and each trip has its own cost of travel. To color a map to describe travel cost, it seems to be necessary to put all the colors everywhere.

Boundaries are equivalent to the edges of colored areas. Every point is on one side or the other, but never both. Map boundaries are what is called a strict function. They assign exactly one element of their range (one bounded region) to each element of their domain (one point on the map).

One-to-One Correspondence

For boundaries to describe a characteristic of the Earth's surface, that characteristic must itself be a strict function, as, for instance, elevation is. Contours of elevation can be

⁷⁷ The number of bends does not affect the length of the route. Turns may slow a vehicle, but they do not add distance. If all its segments are within the limit of detour, a zigzag path is no longer than a simple isosceles path at the same limit of detour.

defined because normally there is only one elevation at each point on the surface. Occasionally, as at “the dreadful cliff that beetles oer his base into the sea,” there is one elevation on the surface where Hamlet stands, a second at the sea and a third under the overhanging face. Overhanging cliffs, tunnels, and caves cannot be described by contour maps, because in these situations, elevation is not a strict function. Though conventional symbols are some help, there is no complete escape from this fundamental limit of the map.

We will call this the problem of establishing a one-to-one correspondence between trips and points.⁷⁸ Because there are many trips from each point, there is no one-to-one correspondence between trips and points. The function that assigns one value of travel cost to one trip assigns many different values to one point. A contour of travel cost cannot be constructed if one point has to be on both sides of it.

One way to solve this problem is to assume a single origin point. Krauske’s isochronic map (described by Brunhes⁷⁹) is a contour map of travel cost. Krauske describes travel time from Berlin. By assuming one origin for all trips, Krauske defines a set of trips in which there is one trip to each point on the map. He uses the minimum travel time read from railroad and steamship schedules to define one value of travel cost for each trip. He can then construct isolines around areas of the world that could be reached in a similar length of time when starting from Berlin. This method works only when a single origin point is assumed. Krauske used Berlin, J. G. Bartholomew⁸⁰ used London. Travel time from Rome to Madrid, for instance, does not appear on either map. Although there is a single minimum travel time from Berlin to any other point in the world, each point has other travel times from other places.

The problem of mapping travel cost is essentially the same as that faced by an architect drawing a plan of a three dimensional house on a two dimensional sheet of paper. He can,

⁷⁸ To call this establishing a one-to-one correspondence means using an old definition of the term. Twentieth century mathematics accepts as a theorem that there is a one-to-one correspondence between the points in a line and the points in a plane simply because both are infinite, and that the number of dimensions makes no difference. Unfortunately, this approach makes the problem disappear without solving it. It is a real problem and has a real solution that can be described by a simple and intuitively obvious analysis of location co-ordinates.

⁷⁹ Brunhes, p. 572.

⁸⁰ Brunhes, p. 572.

with a pencil, assign only one characteristic to each point of the plan, but that point corresponds to a point with shingles on the roof and to points with other characteristics on each floor. The architect solves this problem by slicing through the volume of the house to draw a flat plan for a single level.

Location Co-Ordinates

This solution can be described as a manipulation of location co-ordinates. A point in a volume has three location co-ordinates, x , y , and z . By holding the z co-ordinate constant four feet above the floor, an architect defines a two-dimensional subset of the three dimensional set of points (figure 23). This subset is in one-to-one correspondence with the points of the two dimensional plan. When a point is marked on the plan it defines x and y co-ordinates, and z has been assumed. One point on the plan defines one point in the house and one color goes on the plan.

Krauske's solution to the problem of establishing a one-to-one correspondence between trips and points on the isochronic map can also be described as a manipulation of location co-ordinates. A trip has four co-ordinates: two define the origin and two the destination. The set of all trips is nothing more than the set of all combinations of four location co-ordinates on the map. By assuming the two co-ordinates of Berlin, Krauske slices through the set of all trips and defines a subset that is in one-to-one correspondence with the points of a two dimensional map. The two location co-ordinates of a destination identify a unique member of this subset. With one trip for each point on the map, there is one travel cost, and the point is on only one side of a contour of relative travel cost. Unfortunately, to use this solution to describe all trips would take as many separate maps as there are places.

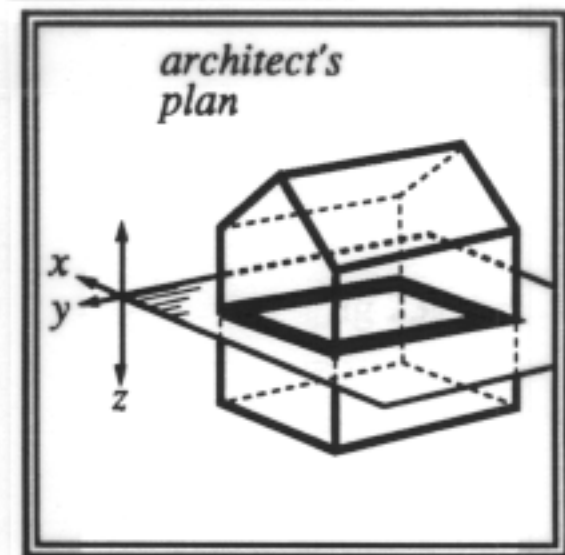


figure 23

Krauske slices through the set of all trips and defines a subset that is in one-to-one correspondence with the points of a two dimensional map. The two location co-ordinates of a destination identify a unique member of this subset. With one trip for each point on the map, there is one travel cost, and the point is on only one side of a contour of relative travel cost. Unfortunately, to use this solution to describe all trips would take as many separate maps as there are places.

Orientation as a Location Co-ordinate

Thünen's derived map of land use demonstrates a more elaborate manipulation of co-ordinates.⁸¹ His economic function contains trip length as a factor, so its value can be defined only for pairs of points. In effect, Thünen's contours are contours of travel cost.

⁸¹ Thünen, *op. cit.*

Thünen establishes a one-to-one correspondence between trips and points by assuming a single destination for all trips, but this is only the beginning. Next he describes the surface of the map by a system of polar co-ordinates with its origin on the market city. In this system, one co-ordinate is the orientation of the trip and the other is distance from the market, which is the same as trip length.

By further assuming a single orientation, Thünen slices the set of trips one more time and defines a one-dimensional sub-set. This set of trips is in one-to-one correspondence with the ordinate of a graph. Each point on this ordinate line corresponds to one trip in the defined set of trips. This allows Thünen to construct a two-dimensional graph with location on one axis and net profit on the other (see page 36). The point on the ordinate defined by the intersection of graph curves is a solution to the profit equation expressed as a location co-ordinate.

When Thünen rotates the graph ordinate around the origin to generate a plane, he reconstructs a larger set of trips, somewhat the way the architect reconstructs the volume of the house. When the architect draws a series of plans for different levels of the house, he removes the assumption of a constant z co-ordinate and substitutes a series of separate assumptions. Thünen entirely removes the assumed location co-ordinate.

The set of trips with a common destination, a common orientation, and all lengths defines a line, Thünen's ordinate. The set of all trips with a common destination, all orientations, and all lengths defines the plane of the map. The set of all trips with a common destination, all orientations, and one trip length defines a circle, Thünen's land use boundary. By this manipulation of location co-ordinates, Thünen describes a difference between trips by constructing a boundary between areas. His boundary separates points, but describes a difference between pairs of points, or trips.

The Set of Trips in One Orientation

Constructing contours of relative travel cost that describe the shadow effect of rivers requires a solution to the same problem Thünen solves, but it needs a different slicing of the set of all trips. The simple approach of assuming one destination will not solve it. Though it is true that all points in a natural route are origins of natural trips to any destination that is in that route, most of these trips are too long to be measured by the rhombus of direct paths. Just as when trying to measure a path that crosses the map, the

rhombus would be too wide. The set of short paths contained in a natural route are not contained in the set of trips with any single destination.

An architect faces a similar problem when he must draw a facade. The set of points that describe the shape of a window are not contained in the set of points four feet from the floor. To show the relations between these points, the architect slices the house vertically and draws an elevation. Holding the x co-ordinate constant

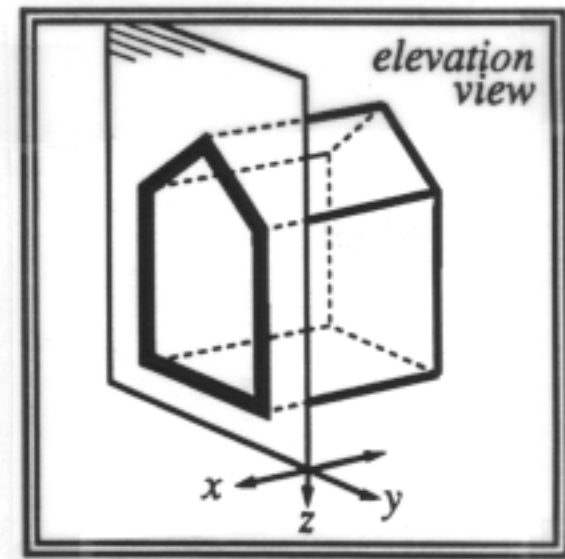


figure 24

instead of the z , he defines a vertical plane on which he can map relations between y and z in two dimensions (figure 24). A one-to-one correspondence can be established between the three dimensional house and the two dimensional plan by assuming any one of the three co-ordinates.

The set of all trips can also be sliced in several different directions. Assuming any one of the four location co-ordinates defines a subset that can be described by only three co-ordinates. Assuming any two co-ordinates defines a set that is in one-to-one correspondence with the map. With one trip for each point, boundaries can describe trips.

In the polar co-ordinate system with its origin on the origin of the trip, the orientation of the trip is one of the location co-ordinates. The set of trips in a direct route all have approximately the same orientation. Assuming a constant orientation slices the set of trips so that all the trips that accumulate in a direct route stay together in one slice. Contours of travel cost can be constructed that describe trips in one orientation.

A Boundary That Describes Trips

To construct a contour of travel cost with origins of easy trips on one side, we begin by assuming an orientation. We will call it the **standard orientation**. We next assume a rectangular co-ordinate system with the y axis in the standard orientation. If we assume one value of x and one value of y , we have defined one point which is the origin of a trip. If we now assume a trip length, we have defined the four location co-ordinates of one trip. The standard orientation and the trip length define the destination of the trip in the polar co-ordinate system whose origin is the origin of the trip. This trip can now be plotted on the map.

For the purpose of demonstration, we will temporarily assume that an easy trip is one for which the straight path (rather than some direct path) does not cross a river. The straight path of the trip we have defined can easily be plotted, and if it does not cross a river, this is an easy trip.

The set of all trips (all combinations of four location co-ordinates) can be reconstructed, the way Thünen does it, by removing, one by one, the four assumptions with which we have defined one trip. Each removal establishes a new defined set of trips (sets 1, 2, 3, & 4).

Removing the assumption that the x co-ordinate of the trip origin is given defines a set of trips, all of the same length and parallel to the standard orientation (set 1). Their origin

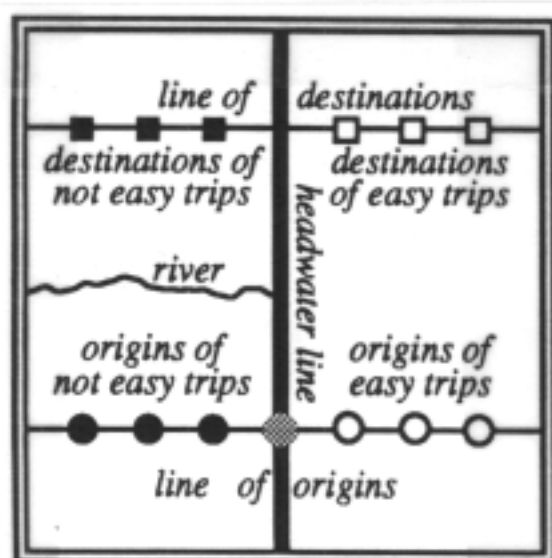


figure 25

points form the line with the given y co-ordinate and all possible values of x . We will call this the **line of origins**. The destinations of these trips form a **line of destinations** parallel to, and one trip length away from, the line of origins.

If a river lies between the line of origins and the line of destinations, trips above the headwater are easy, and trips below it are not (figure 25). A line can be projected in the standard orientation through the headwater of the river. We will call this the **headwater line**. The point at which the headwater line intersects the line of origins defines a boundary point. When we arbitrarily call the side with the river inside, all easy trips in the defined set of trips (set 1) start from the line of origins outside the boundary point.

Assuming a different trip length moves the line of destinations, but the line of origins and the headwater line remain the same. As long as the river is between the line of origins and the line of destinations, the boundary point remains the same. All points that are origins of easy trips in the defined set of trips (set 2) lie outside the boundary point. When the river is not between the two lines, the boundary point becomes meaningless, and all trips in the defined set are easy.

If the assumption that trip length is given is removed entirely, the defined set of trips is enlarged to include trips of all lengths that start from the line of origins and run in the standard orientation (set 3). The line of origins and the boundary point remain constant.

If we define trips in the opposite direction by negative trip lengths, all possible destinations are described. All easy trips in the defined set (set 3) start from the line of origins outside the boundary point, unless both endpoints are on the same side of the river (figure 26).

So far the line of origins is constant. All trips in the defined set (set 3) start from points with the assumed y co-ordinate. If this assumption is also removed, the line of origins moves to generate the plane of the map: the set of all points with all values of x and all values of y . The defined set of trips now contains all trips in the standard orientation (set 4). As the line of origins moves, the boundary point generates a boundary line identical to the headwater line. All points that are origins of easy trips in the standard orientation lie outside this boundary unless the trip is too short to reach the river.

A trip with its origin and destination on the same side of the river is always easy, but such trips cannot form endless routes. An endless route must contain some paths that reach past the river, either by crossing it, or by going around its head. Easy trips that reach past the river all have their origins outside the headwater line if they run in the standard orientation.

To completely reconstruct the set of all trips, the assumption of a given standard orientation must also be removed, but when the standard orientation changes, the headwater line rotates and no longer divides the map. The headwater line describes only those trips whose orientation is the same as the standard orientation. Given this limitation, the headwater line finds all straight paths that avoid crossing a river.

A Boundary That Describes Direct Paths

The headwater line describes only straight paths, but natural paths may bend. Our temporary definition of an easy trip as a pair of points between which the straight path crosses no river is too restrictive. It can be replaced by a more realistic definition by incorporating the rhombus of direct paths into the construction of the boundary. From now on, we will define an **easy trip** as a pair of points connected by a direct path (rather than a straight path) that crosses no river.

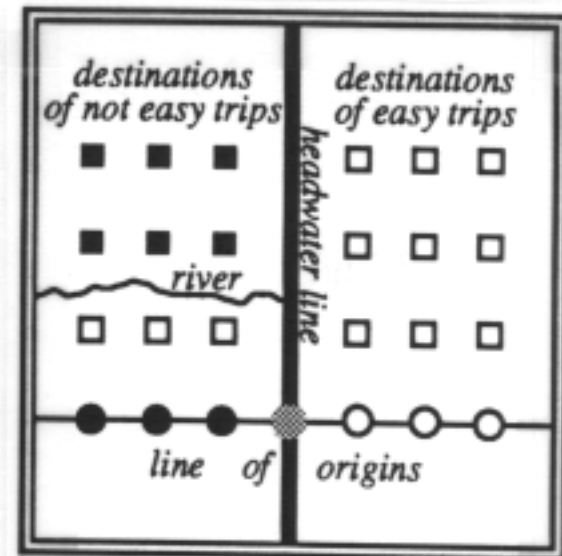


figure 26

We have defined a direct path as one that remains within the rhombus of direct paths. If no river enters the rhombus, or a river enters but does not cross it, some direct path avoids the river and the trip is an easy trip.

An angle can be constructed with its vertex on the headwater point and with each leg at thirty degrees to the standard orientation. This curve encloses the river like a broad arrow head pointing up stream (figure 27). We will call this figure the **headwater curve**. It can

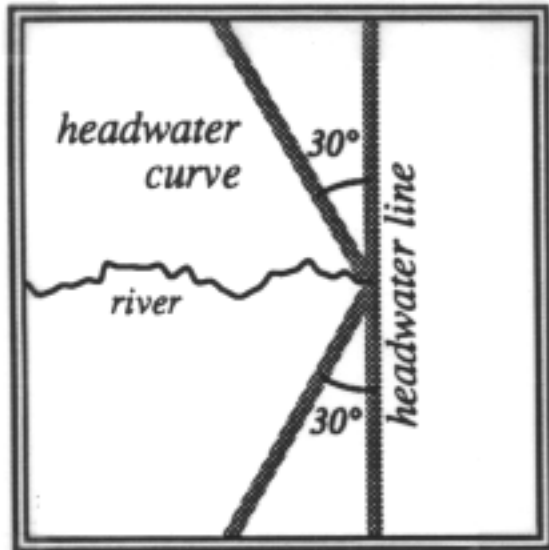


figure 27

be shown that all easy trips past the river have their origins outside the headwater curve.

To do this, we will again define a set of trips whose origins are in a line perpendicular to the standard orientation (set 1). We will assume a trip length that puts the headwater halfway between the line of origins and the line of destinations. The headwater curve intersects these two lines at the endpoints of a trip in the standard orientation. The headwater curve coincides

with an isosceles path for this trip. It also coincides with two sides of the rhombus constructed on those same points. All paths contained in the rhombus cross the river and this is not an easy trip.

If a similar rhombus is constructed from any point on the line of origins inside the headwater curve (figure 28), all the paths it contains must cross the river. If the rhombus is constructed on an origin outside the headwater curve, it contains some path that avoids the river.

Where the headwater curve intersects the line of origins is a boundary point. Points in the line of origins outside this boundary point are the origins of easy trips in the defined set of trips (set 1). Points inside the boundary point are not.

When the defined set of trips is enlarged to include trips of all lengths (set 3) and the line of destinations moves, the figure (figure 28) remains geometrically similar only if the *y* coordinate of the origins changes at the

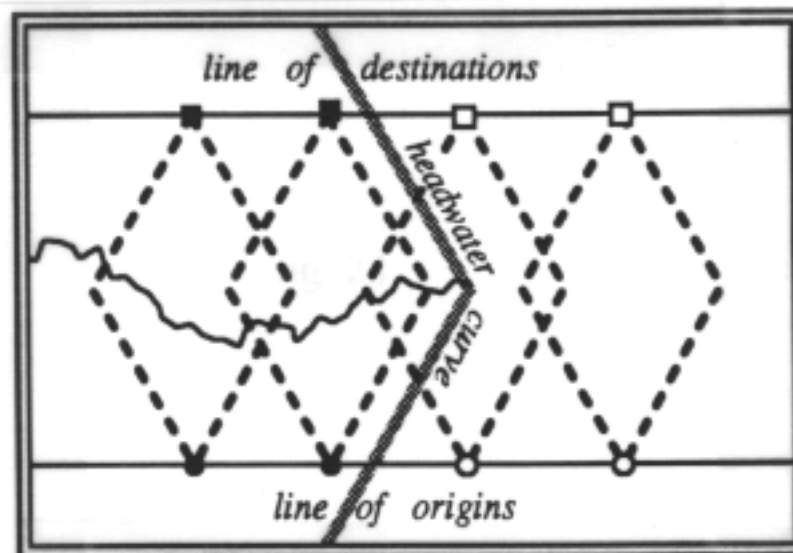


figure 28

same time, so that the river always remains halfway between the line of origins and the line of destinations. The boundary point describes only symmetrical cases. In asymmetrical cases, where the river is not halfway between the lines of endpoints, the headwater curve

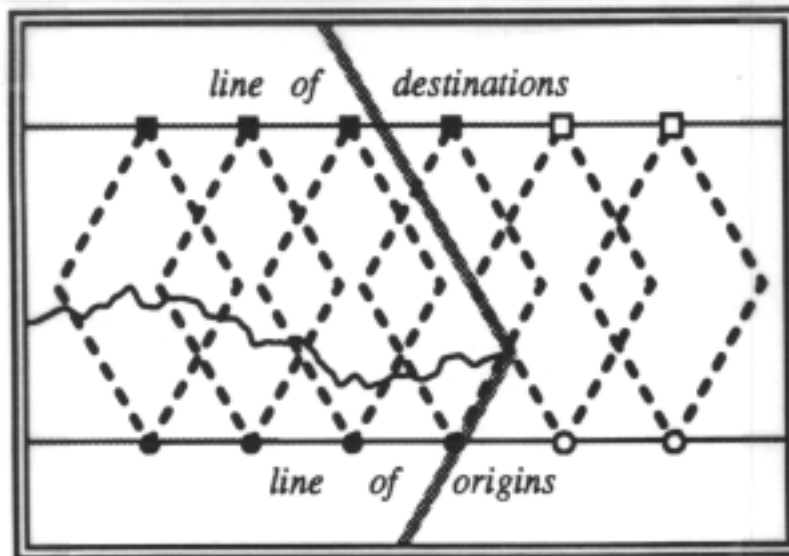


figure 29

describes a path between them, but the endpoints of that path are not a trip in the standard orientation.

In an asymmetrical case, the rhombus can be constructed so the river just crosses it and its upstream side passes through the headwater point. The trip it describes and all trips down river are not easy. All easy trips in the defined set

(set 1) are on the upstream side of this rhombus (figure 29). Since part of the side of this rhombus coincides with the headwater curve, one endpoint of this trip is on the headwater curve. The other endpoint is outside the curve. A rhombus constructed farther upstream has both endpoints outside. Therefore, it remains true in all cases, whether symmetrical or not, that if an easy trip reaches past the river in the standard orientation, its origin lies outside the headwater curve.

The headwater curve describes the shadow effect of one river, how a river affects the cost of travel on the land leading up to and beyond it. If no other river is near, the headwater curve is a contour of relative travel cost, showing that travel in the assumed orientation is easier outside it.

Routes as Bounded Areas

The headwater curve can be used to find ridge routes. As a boundary in the map, it defines a difference between points. Because of its relation to the rhombus of direct paths, it describes a difference between trips. We will show how it can be interpreted to describe a difference between routes.

We will define a **through route** as a direct route in the standard orientation that crosses the map from edge to edge. A ridge route is a direct route that crosses the map. If it also has a general orientation the same as the standard orientation, it is a through

route as here defined. The definition of a through route incorporates the assumption of a standard orientation throughout the following argument.

A through route has two of the characteristics of a ridge route. It is direct, and it has the maximum possible extension. If the edges of the map are parallel, all through routes create the same benefit and have the same base cost. When it is assumed that detour cost is zero on any direct path, it is zero on any through route. Only the cost of crossing rivers can affect the relative costs and benefits of through routes. We have shown that crossing cost is equivalent to the number of rivers crossed (see page 40). Therefore, if there is a ridge route in the standard orientation, it is the through route with the minimum number of river crossings. To find ridge routes is to find where direct routes avoid rivers.

The headwater curve describes a difference between through routes. A through route that enters the area inside a headwater curve must cross the river on which that curve is constructed. This can be proven from the definitions and conclusions we have reached.⁸² The headwater curve for one river divides the map into two areas. The area outside the curve contains through routes that avoid the river, the area inside the curve does not.

A Boundary That Finds Ridge Routes

Headwater curves outline route areas that define the locations of ridge routes. We will consider first a special case that we will call the **perfect ridge route**. A perfect ridge route is a through route that crosses no rivers. Since a through route that passes inside a headwater curve must cross a river, a perfect ridge route must remain outside all headwater curves.

If headwater curves are constructed on a map that contains two rivers, the curves may point toward each other and define an hourglass shape between them (figure 30). The

⁸² This proof is as follows: If a point is in a route, it is the origin of a route segment to every other point in the route. Since a through route crosses the map, it contains points on both sides of every river in the map. Therefore, if a point is in a through route it is the origin of a route segment to some point that is beyond the river. Since a through route is direct, all route segments are direct. Therefore, if a point is in a through route, it is the origin of a route segment that is a direct path to a point beyond the given river. A headwater curve can be constructed on this river. If a point is inside the headwater curve, all direct paths from this point to points beyond the river must cross the river. Therefore, if a point is inside the headwater curve, and it is in a through route, some segment of the through route crosses the river. This is the same as to say that if a through route enters the area inside the headwater curve, it crosses the river on which the curve is constructed.



figure 30

area outside both headwater curves contains through routes that avoid both rivers. A through route that leaves this area must cross at least one river.

Headwater curves are here boundaries of a route area. The route area fans out near the edges of the map, but part of it narrowly defines the location of a perfect ridge route.

When headwater curves are constructed on rivers that flow the same way, they will point the other way and intersect each other.

When they do, they unite to form a continuous boundary (figure 31). This boundary is a contour of the natural route function. A road will create more benefit at less cost if it remains outside this boundary.

When headwater curves are constructed on all the rivers of a large map, it may happen that contours of travel cost will define a long narrow area that crosses the map outside all headwater curves. This area contains a perfect ridge route across the map.



figure 31

A route area may contain a route with complex alignment. The area between contours may bend back and forth and so allow variation in the local orientation of the route (figure 32). As long as the route area remains continuous, however, it contains some through route of which no segment is beyond the thirty degree limit of detour.

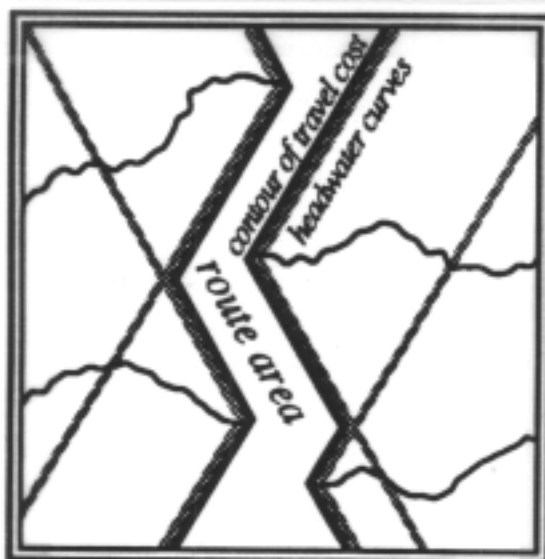


figure 32

Headwater curves show where no direct route avoids rivers. If the headwaters of two rivers reach past each other, the headwater curves point in opposite directions and may intersect (figure 33). The area outside both curves is not continuous and cannot contain a through route. There is a route that avoids both rivers by following the watershed, but it contains a segment more than thirty degrees from the standard orientation. This route is not direct. Headwater curves measure the orientation of the

watershed to test whether it can hold a direct route in the standard orientation.

Contours of relative travel cost constructed from headwater curves test for all the characteristics of ridge routes. If a route is endless, the route area must extend across the map. To contain a continuous route, the route area must be continuous. If the through route is direct and avoids all rivers, it must be outside all headwater curves. Route areas between contours of travel cost find some ridge routes. If a route exists, in the standard orientation, that fits the definition of a perfect ridge route it will appear in a route area.

A Contour Map of Travel Cost

So far, headwater curves can find only perfect ridge routes, the special case of ridge routes that avoid rivers absolutely. Perfect ridge routes are common only within small regions. Ridge routes across large regions cross rivers to join up in a direct line the perfect ridge routes of smaller regions. A **normal ridge route** is a through route that crosses a local minimum number of rivers. To find these routes the local minimum must be defined by comparing all routes near.

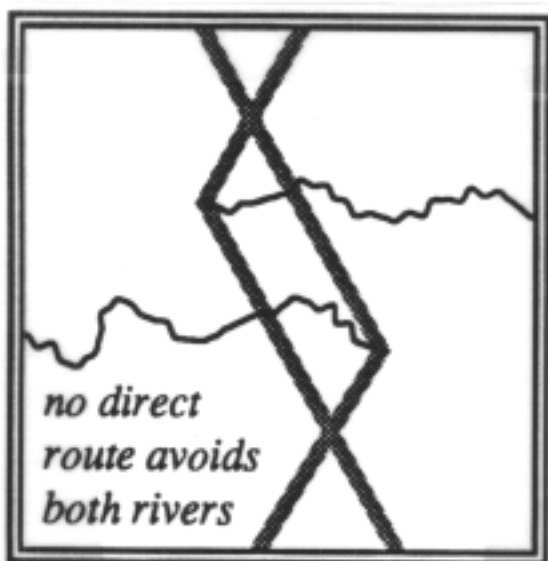


figure 33

be defined by comparing all routes near.

The number of rivers that a through route must cross can be identified for any area of the map. In the example where two headwater curves point in opposite directions and intersect, they divide the map into five subdivisions (figure 33).

The east subdivision is outside the west headwater curve. It contains through routes that avoid the west river, but all routes in this area pass inside the east headwater curve and must cross the east river. The similar area on the west side of the map contains only routes that cross the west river. Any through route that enters either of these two areas must cross at least one river.

The skewed diamond in the middle of the map is inside both headwater curves. Because it is inside the west curve, through routes that enter this area must cross the west river. Because it is also inside the east headwater curve, any through route that enters this area must cross both rivers. The north and south subdivisions are outside both headwater curves, but they do not form a continuous area across the map and cannot contain a through route.

A route area can be assembled that includes the north area by adding the east area to it. This area is outside the west headwater curve, but through routes in it all cross the east curve. Through routes can cross a headwater curve only if they also cross the river on which the curve is constructed. The assembled area contains through routes that cross one river.

Other route areas can be assembled from combinations of the four areas around the edges of the map. If the areas that are combined are contiguous, the assembled area is continuous. All through routes in these combinations of contiguous areas must cross at least one river.

Road Value Number

Each of the five subdivisions of this map can be given a number that is the least number of crossings in any through route entering that area (figure 34). This we will call the **road value number**. In all regions outside the center, the road value number is one. In the skewed diamond in the middle, where all through routes cross both rivers, the road value number is two.

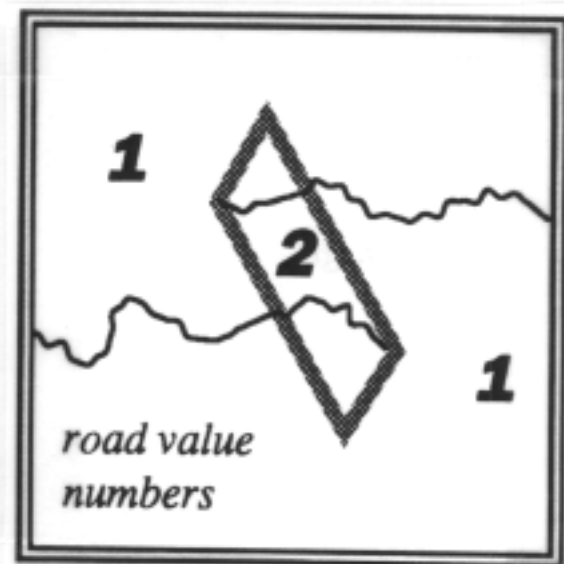


figure 34

The road value number describes the range of the natural route function (see page 34). It is an inverse

scale that defines the value of a point for use as a road. A road will create the most benefit at the least cost if it remains in the area with the lowest road value number. When headwater curves separate areas with different road value numbers, they are contours of the natural route function. The area on one side contains better routes than the area on the other side. When headwater curves run between areas with the same road value number, they are meaningless.⁸³

Confluence Curves

The number of crossings in a through route changes not only at the headwaters of rivers, but also at their forks. A road that crosses above the fork must cross two branches of the stream, while below the fork there is only one crossing. This difference can be

⁸³ There may be exceptions where headwater curves with the same road value number on both sides distinguish variant routes (see page 65).

described by a boundary similar to the headwater curve. Lines projected from the river junction at thirty degrees to the standard orientation and enclosing the double stream define a **confluence curve** (figure 35). This curve is analogous to the headwater curve. Through routes that enter the area inside the confluence curve must cross both branches of the river, while through routes that remain outside need cross only one.

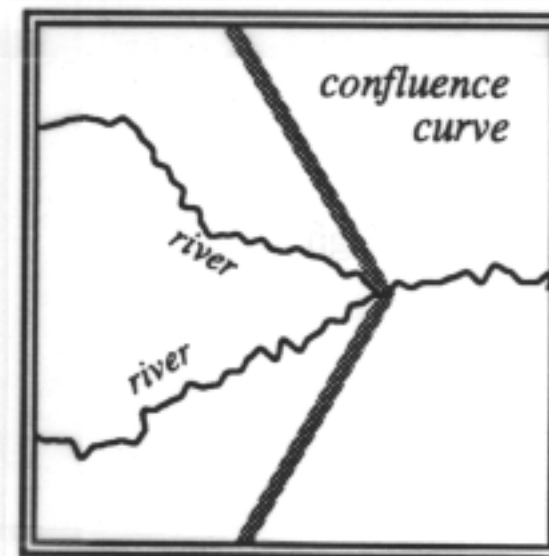


figure 35

Headwater curves and confluence curves can be projected from every headwater and fork of a river system.⁸⁴ Road value numbers can be assigned to every point on the land. No matter how complex the map, the minimum number of river crossings is unique for every area. The boundaries define a very large number of subdivisions, but many can be assembled into larger regions. The result is a contour map of travel cost that assigns to every point on the land a measure of the relative cost of overland travel in the standard orientation. Ridge routes can be read from this map.

Areas That Contain Routes

The contour map of travel cost shows areas that contain ridge routes. When road value numbers are assigned to the entire map of a region, the lowest numbers apply to continuous strips across the map. These areas contain through routes with the local minimum number of river crossings. We will call them **primary route areas**. They

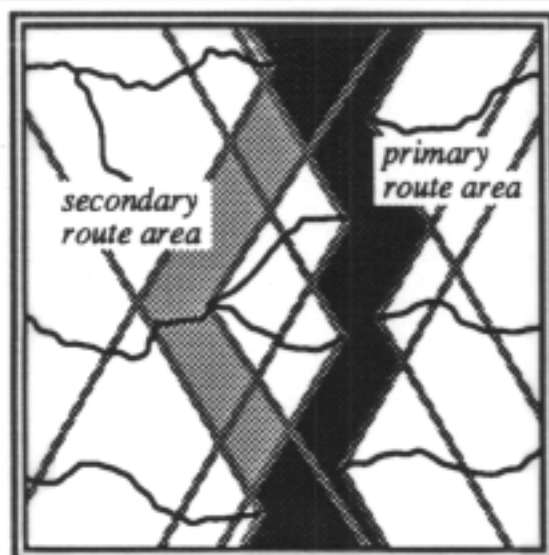


figure 36

contain all routes that fit the strict definition of a ridge route. No fixed number of crossings defines a primary route area. A road value number that is a local minimum in one place may be on a shadow area in another. A number is a local minimum when no contiguous area has a lower number.

The highest numbers identify shadow areas, generally diamond shaped, where all through routes must cross

⁸⁴ Curves at elbows may also be necessary.

numerous rivers. These areas are not continuous across the map, but are surrounded by areas that contain better routes.

Intermediate numbers also appear. Each primary route area has a continuous border of trivial deviations that cross one extra river. Other areas with numbers near the minimum define distinct alternate routes. These we will call **secondary route areas** (figure 36). They do not form continuous areas across the map. They form branches, loops, and crossovers between primary routes. To reach the edge of the map, the routes they contain must join the primary routes. Though in the next chapter, we will see how secondary route areas may, under certain circumstances, be the best locations for roads, most early roads in humid regions of slight relief should be expected to follow the natural routes described by primary route areas.

All Routes Described

The contour map of travel cost describes all paths in one orientation. By defining the least cost overland path, we have already reduced the set of all paths to a set that includes only one path for each trip. The set of all trips is all combinations of four location coordinates. Points on the map define two of these co-ordinates, and the contours describe all points. The third co-ordinate is trip length. This is described by the condition that trips must reach past a river, and is otherwise an insignificant difference. The fourth co-ordinate is the assumed standard orientation. The contour map of travel cost describes all combinations of four location co-ordinates in which one co-ordinate is the standard orientation.

To reconstruct the set of all trips and completely describe the set of all paths, the assumption of a given orientation must be removed, but when this is done, the headwater curves rotate and the map can no longer be constructed.

The contour map of travel cost can describe only one orientation because it is the nature of travel cost to be specific to one orientation. A ridge that makes a good road in its long axis, makes a bad road in the cross direction in which it leads out of one river and into another. If all orientations are considered, the same point both does and does not have the characteristics of a natural route, and there cannot be one color at each point of the map. One map cannot describe all orientations but the set of all trips can be

reconstructed by making a small number of separate maps for separate orientations the way an architect makes separate plans for every floor.

Though headwater curves are constructed with an assumed standard orientation, this assumption does not completely determine the orientation of the routes that are found. There is a tolerance in the limit of detour that allows variation in the local orientation of a route. This same tolerance also allows variation in its general orientation. If a route is a straight line it can be up to thirty degrees from the standard orientation and all its segments will still be within thirty degrees of the standard. It will appear as a direct route on a contour map of travel cost constructed for any orientation within thirty degrees of its own.

Routes that bend must be closer to the standard to be found. The contour map of travel cost compares all possible alignments within a range of orientations near the standard, but it tests for a limit of detour that varies slightly with the orientation of the route. A route that detours by a full thirty degrees will appear only if its general trend is exactly the standard. Headwater curves are less tolerant of bends in a route that is farther from the standard orientation.

The definition of "direct" by a thirty degree limit of detour is generous. Fifteen degrees is actually a better description of the observable pattern of historic roads. We can tighten the definition of a direct route to be one that remains within fifteen degrees of a straight line. With this new definition, all segments of a direct route are within thirty degrees of a standard orientation that is itself within fifteen degrees of the general orientation of the route. Contours of travel cost constructed on thirty degree angles will find all direct routes in a thirty degree range of orientations. If six maps are constructed with standard orientations thirty degrees apart, all orientations are covered.⁸⁵ Any endless route that avoids rivers and is direct by this narrower definition will appear as a route area on at least one of these maps.

A Model of Ridge Routes Completed

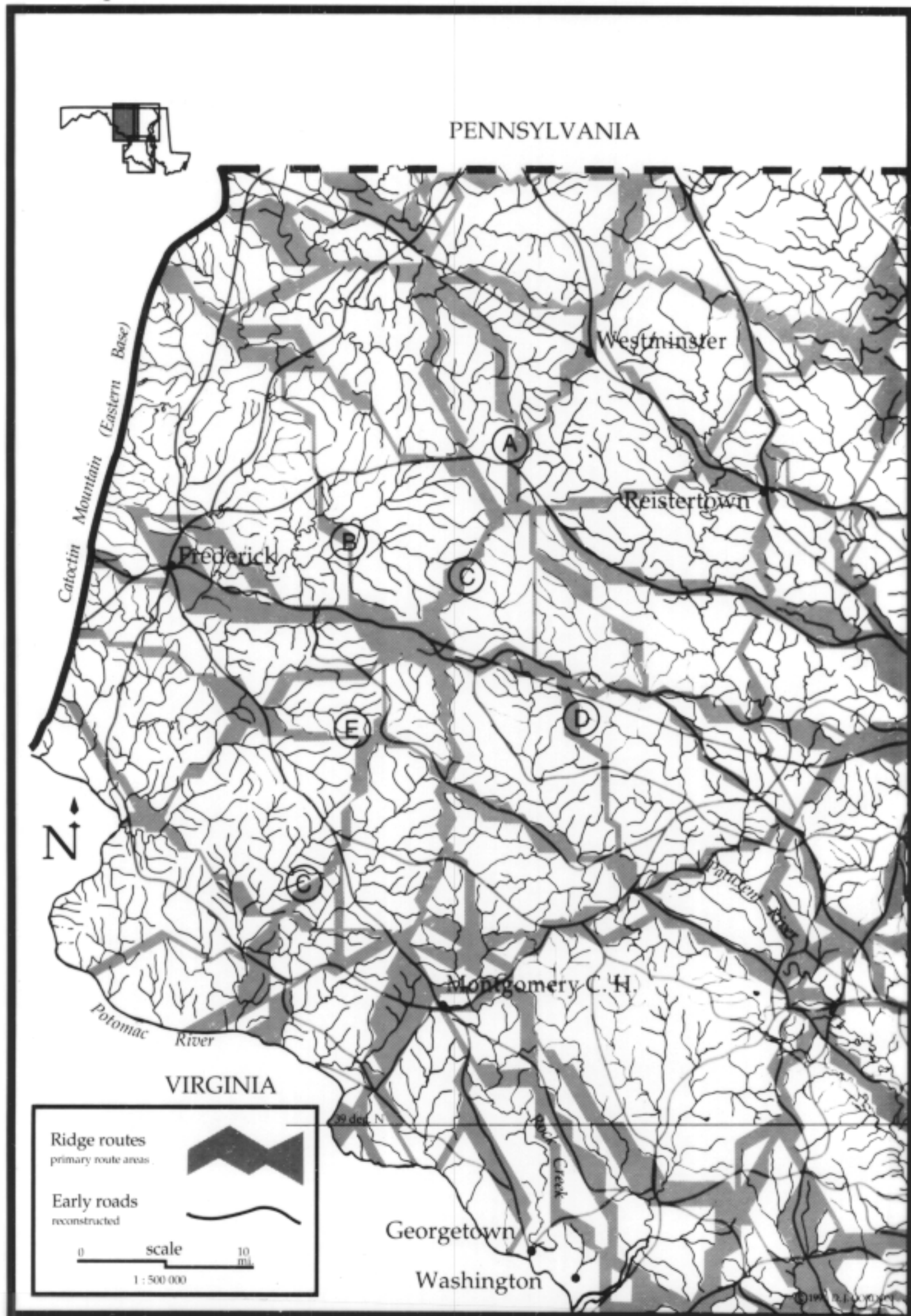
The contour map of travel cost is a predictive model of the hypothesis that roads follow ridge routes. It examines all possible routes to select natural routes. Headwater curves

⁸⁵ It may be noticed that combinations of trips in different orientations have not been examined, but these do not form direct routes.

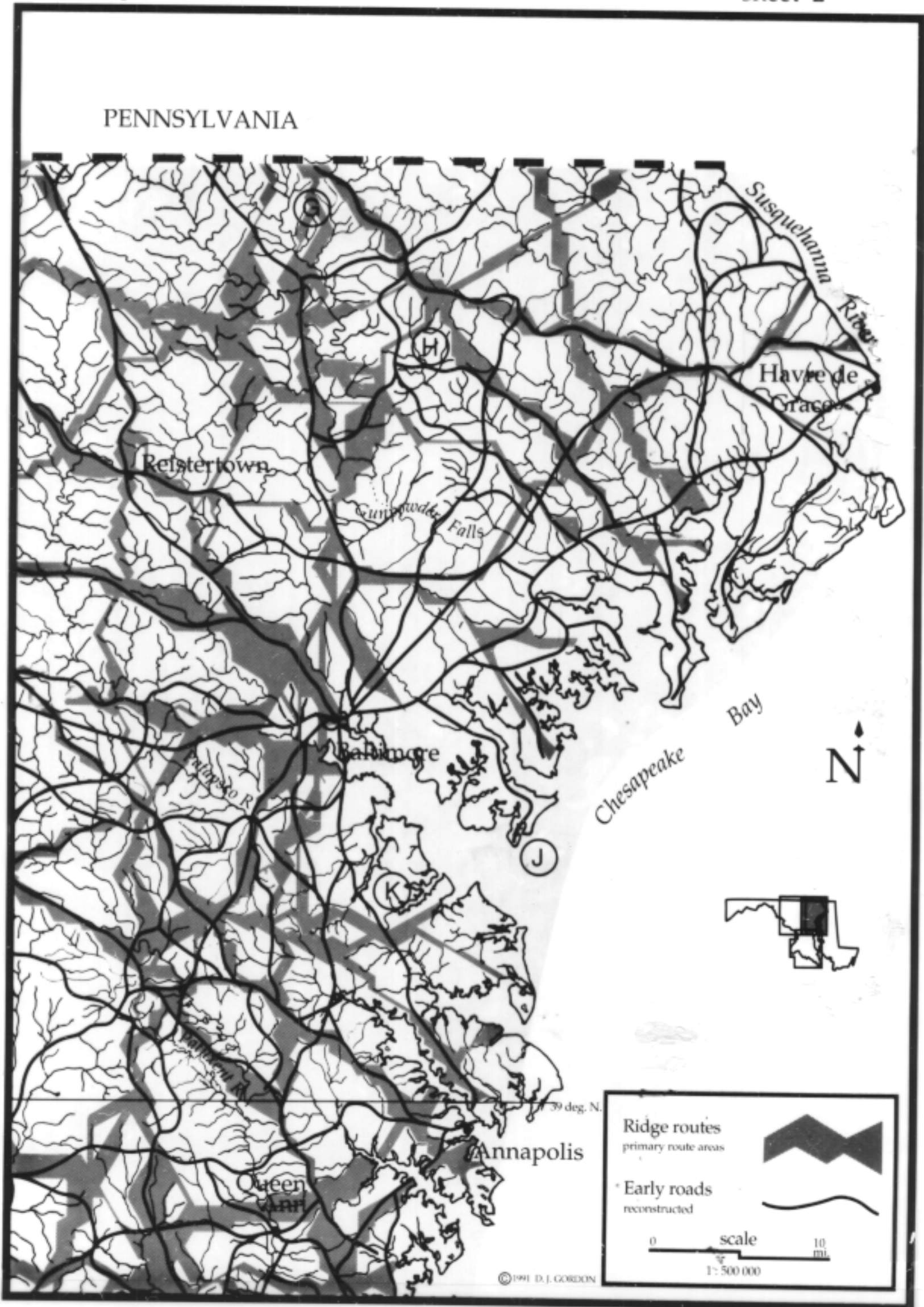
measure the spatial relationships between rivers to identify ridges that form direct routes. These mechanically constructed boundaries identify routes on an objective and rigorous basis. This model finds all routes that extend across the map, remain within fifteen degrees of a straight line, and cross a local minimum number of rivers. If the hypothesis of ridge roads is valid, this logical condition is equivalent to the practical condition that a road will create the most benefit at the least cost. Headwater curves solve the physical equation, work equals distance times resistance, that describes the cost of overland travel. Route areas across the map solve the economic equation of costs and benefits. The model describes a physical and economic connection through which terrain could act on roads. It predicts where roads would be if this connection acted alone.

A map of ridge routes can be compiled by extracting primary route areas from the separate maps constructed on six orientations. This map leaves out secondary routes, but it can show routes in all orientations. It defines the alignment of routes with considerable precision. This logically derived map can be compared to a map of roads to see whether roads follow ridge routes.

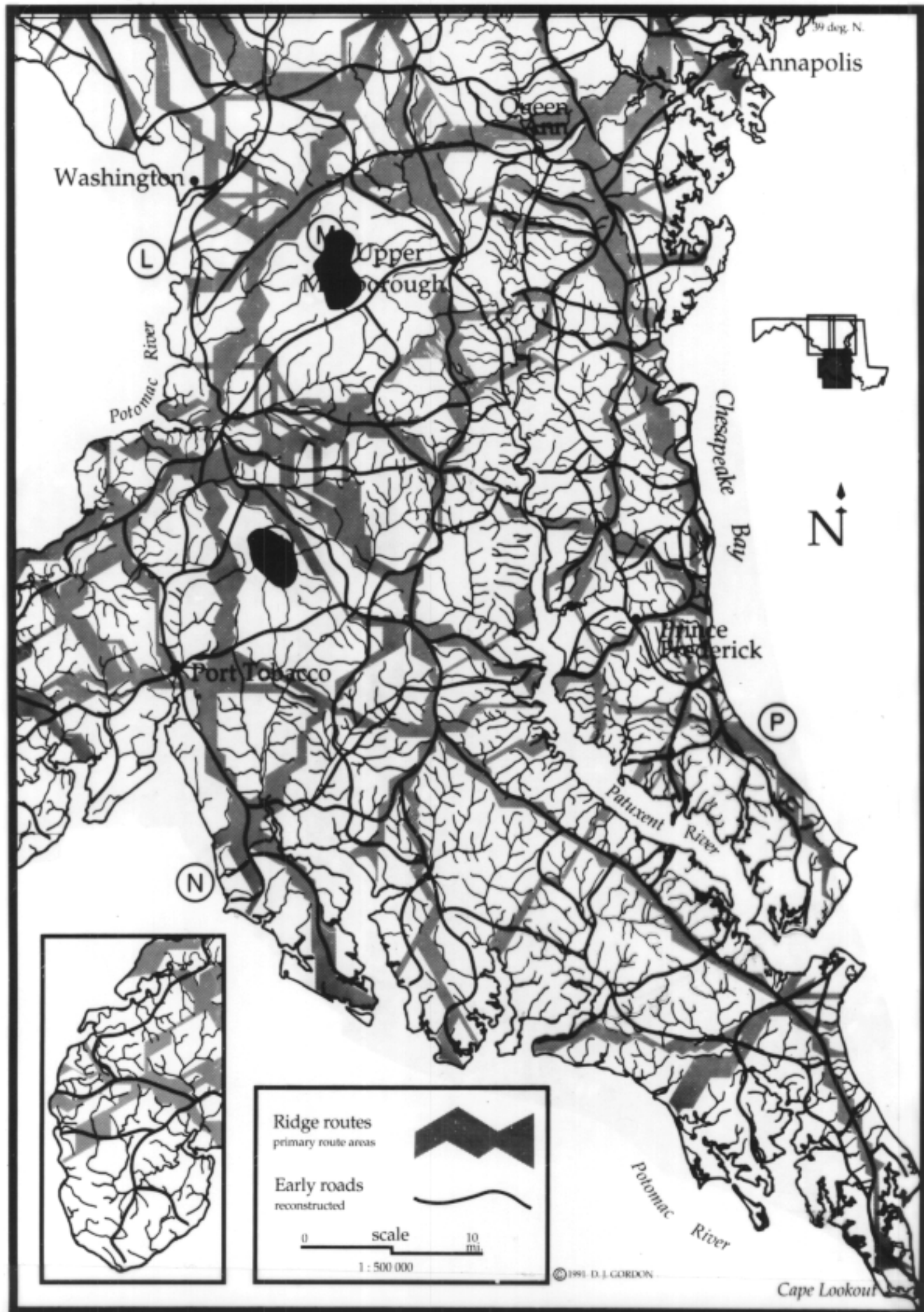
CENTRAL MARYLAND: RIDGE ROUTES AND EARLY ROADS
map one sheet 1



CENTRAL MARYLAND: RIDGE ROUTES AND EARLY ROADS
map one sheet 2



CENTRAL MARYLAND: RIDGE ROUTES AND EARLY ROADS
map one sheet 3



Chapter 4

Central Maryland: a Case Study

The hypothesis of ridge routes suggests that early roads in humid regions of slight relief should follow endless, direct routes that avoid rivers. A map of ridge routes derived by the method described in the preceding chapter can test whether, in fact, they do. We have chosen central Maryland for this purpose. This region is small enough to be practical and fits the normal pattern of humid regions of slight relief. It is a region of low rolling hills with a maximum elevation of 1000 feet. The coastal plain here merges with the piedmont almost imperceptibly. Its drainage is mature enough that there are no large areas of swamp. It is naturally forested with the typical forest soils.

Central Maryland is a clearly defined region, bounded on the east by Chesapeake Bay, on the south by the Potomac River, and on the west by a clear straight line of mountains, a continuation of Virginia's Blue Ridge. Only the northern boundary, the Mason-Dixon Line, is topographically arbitrary. Its area is about 5000 square miles. With forty inches of precipitation, central Maryland is a humid region. It is the kind of region to which the hypothesis of ridge roads applies and resembles the regions where obvious ridge roads can be found. It also contains enough early roads to support generalizations.

Of this region, we have constructed six contour maps of travel cost as described in Chapter Three. A simplified method of constructing these maps, tracing a grid and starting from one edge, is described in Appendix B. From these six maps we have extracted the primary route areas. These form a map of ridge routes that predicts the locations of early roads. This map (with a modification described below [page 75]) is included in Map One.

A Map of Early Roads

To compare the predictions to the facts, this map must be compared to a map of early roads. The availability of an excellent map showing the early roads is, in fact, an important reason for choosing Maryland for a case study. The first road map of Maryland was published by Dennis Griffith in 1794.⁸⁶ Because Maryland is small, Griffith's map is at a larger scale than most American maps of the same period. Because Maryland was already well populated, the map shows a large number of roads. At this early date, it can be reasonably assumed that the roads it shows are primitive roads made of nothing but the natural surface of the land. They should be expected to match the map of ridge routes.

Unfortunately, Griffith's map does not accurately define the exact positions of the roads it shows. The map is not based on a trigonometric survey. Though geodesy had been invented, the ordinary survey of the 1790's was by laying a measuring chain on the road and taking compass bearings.⁸⁷ This method is accurate over short distances, but it tends to accumulate error. Consequently, Griffith's map has serious distortions, and it would be meaningless to superimpose it directly on the predicted map.

The true positions of the early roads can be reconstructed by transferring them from the early map onto the modern geodetic base. Griffith's map is clear in its detail. It shows a dense network of drainage that can easily be identified with streams found on the modern map. In most places that it shows roads, modern roads can be found in the same positions relative to the rivers and with the same number and direction of bends. By assuming, when a road on the modern map closely resembles a road on Griffith's map, that it is the same road, we have constructed a map of early roads on a modern, planimetrically accurate base. This reconstructed map of early roads is superimposed on the map of ridge routes in Map One.

This is not a study in historical geography and no claim is made to the absolute accuracy of this map, but it provides a reasonably faithful definition of the geographic relationship between roads and terrain for a regional network of early roads.

⁸⁶ Dennis Griffith, "Map of the State of Maryland," (1794 [1795]), in Edward C. Papenfuss and Joseph M. Coale III, eds., *The Hammond - Harwood House Atlas of Historical Maps of Maryland, 1608 - 1908*, (Baltimore: Johns Hopkins Univ. Press, 1982), p. 49 - 52.

⁸⁷ Lloyd A. Brown, "Chapter 9," *The Story of Maps*, (Boston: Little Brown and Co., 1949).

Roads Correctly Predicted

With these two maps superimposed, route areas can be compared to roads. If an historic road runs within a predicted area, it matches the prediction. A road that crosses a route area does not match even though it may run within the area for some distance. If a road leaves the predicted area and re-enters, we count as correctly predicted the mileage of the road that is within the area.

Many roads are correctly predicted. Some are obvious, like the road that runs northwest from Cape Lookout, following the long watershed ridge between the Potomac and the Patuxent. The road on the ridge between the Patuxent and the Patapsco might also have been found easily without a predictive model, as could that on the right bank watershed of the Susquehanna.

It is important that a technique of observation should not miss the obvious, but the more interesting predictions are those not so obvious that run across the grain of the country. The road between Baltimore and Montgomery Court House (now Rockville) is the best case of a route that crosses many rivers and could not be recognized in a simple examination of watersheds, that yet is almost perfectly predicted by the contour map of travel cost. Similar, though less accurately predicted, is the road west from Queen Ann. The road south from Upper Marlborough and the northernmost of the three roads between Baltimore and Frederick are also very accurate predictions of roads that are not merely following basin boundaries.

All told, the map of early roads shows 1678 miles of road in central Maryland. Of these, 643 miles (38.3 percent) match the prediction and follow ridge routes.

This correspondence cannot be random. Though almost thirty percent of the area of the map is contained in one predicted route or another, a road does not match the prediction unless it agrees not just in position, but also in orientation. With six different orientations, the random odds of matching routes to roads in both position and orientation are more like one sixth of thirty percent. Even five percent is an overstatement. This would be the odds if roads and routes were all straight lines, which they are not. The true odds cannot be computed because the cases are not separate and the total number of possible routes is completely undefinable, but the details of the alignment of many roads is a very large number of cases. Many small bends are located exactly. It is obviously

impossible that the match between early roads and predicted ridge routes should arise by random coincidence.

The alternative to coincidence is cause. With the working connection, known to operate, through which ridges could act on roads, with the unquestionable independence of terrain from roads, and with the consistent correlation demonstrated, cause and effect can be inferred. Ridge routes, the natural routes of humid regions of slight relief, cause the location of many early roads in central Maryland.

Roads Not Predicted

There remain many differences between the predicted routes and the early roads. These could represent either failures of the hypothesis or the effects of modelling simplifications. In constructing the model we assumed that rivers are the only resistance to travel, that the limit of detour is constant, and that the interactions of different roads have no effect. Each of these simplifying assumptions excludes recognized forces that can explain many of the roads that are not predicted. Roads not predicted do not contradict the hypothesis if their cases can be explained by modelling simplifications.

Some roads are partially predicted. The map of early roads can be compared, not only to the map of ridge routes which shows only primary route areas, but also to the six contour maps of travel cost which show secondary routes as well. Many of the roads that are not predicted are contained in secondary route areas. These route areas are too numerous to be included on the map of predicted routes or to be considered clear predictions, but they definitely show ridge routes. Whatever other forces affect these roads need explain only a minor difference.

Roads Affected by Other Terrain Features

The assumption that only rivers matter to travel excludes a host of features, many of which can be found on the topographic map. At map letter P the primary predicted route area is crossed by deep gorges and the road follows a secondary route area. At map letter A, there is a range of hills and the road uses a pass.

These are relatively gross features and can easily be identified. A more detailed examination of the land than falls within the limits of this work would no doubt show finer features influencing roads, especially the ledges in the rivers that provide a rock bottom

for a bridge or ford. Thus some roads remain unexplained even though they might be completely determined by terrain.

There is one more kind of terrain feature that deserves special notice. There are two areas (map letters M) at the meetings of watersheds that, by the rigorous prediction, should be the junctions of several roads but which, in fact, have no roads. These areas are elevated but flat. They are picked out by the model as land away from rivers, but they are not ridges. A flat area with no river through it cannot be well-drained land. Modern paved roads, with artificial drainage, can and do pass through these areas. Early roads do not.

Undrained elevations can be incorporated in the model. When added to the map of rivers and treated as places where roads do not go, they change the predicted routes. Primary routes that cross the undrained elevations disappear, and parallel secondary routes become primary routes that correctly predict roads. It is this modified map of ridge routes that is shown on Map One. Though slightly less rigorous, this modification of the model is a truer description of the hypothesis and raises the mileage of correctly predicted roads to 688 (41.0 percent).

Roads Showing Variation in Detour

There are several roads in Maryland that are on watersheds but do not remain direct. They diverge from the straight line by more than thirty degrees, and their routes are automatically rejected by the model. East of Baltimore, North Point Road (map letter J) follows the high ground on a neck of land that juts out into Chesapeake Bay. To stay direct it would have to cross not streams, but wide coves. Clearly, the failure of the model to predict this road demonstrates only that not all rivers are the same. There are similar roads on other necks.

Inland, too, there are some roads on ridges that are not direct. One such road is at map letter E. Most of this road follows a secondary route area, but leaves it to follow the watershed around a sharper than normal bend. It thus avoids not one, but several river crossings. The constant limit of detour does not compare the cost of crossing small rivers against large, or of several rivers against none. Such roads as clearly follow ridges do not contradict the hypothesis of ridge roads. They show only that the constant limit of detour is an oversimplification

One more effect of the limit of detour is not a variation in how much roads bend, but a variation in the limit that is actually tested for. As mentioned above (page 66), if a route has a general orientation the same as the standard, it can diverge from the straight line by a full thirty degrees, but between two standards a route must remain within fifteen degrees of a straight line or it will not appear as a continuous route area. Because the thirty degree limit of detour is generous, this variability in the measured limit of detour will not eliminate any direct routes, but it will sometimes cause a route to be found which is not direct by the narrower definition of direct. The true ridge route will then appear as a secondary route area. This variability in the measured limit of detour may explain some of the roads that follow secondary routes.

Roads Affected by Interconnection

The model isolates routes from their surroundings, but roads are not isolated entities. The model does not allow for the fact that combining and connecting roads together can reduce cost and add benefits.

That roads connect to navigation stands as a fundamental assumption of the model. The shoreline is counted as the edge of the map and stands for the goal of the road. In Maryland, surrounded by bays, nearly all the ridge routes reach water. They define sections of shoreline with superior inland connections.⁸⁸ Within these areas a choice remains of where to meet navigation that can be based on the quality of the anchorage and of the foreshore. These requirements of a port may also cause a road to bend. The best ridge route in Maryland runs down the neck that Annapolis is on, but to reach the harbor there, the road bends out of a direct line.

The roads at Baltimore provide a catalogue of the minor modifications that interconnection can effect. Only two of these are exactly predicted, the road from the west-northwest and that from the north. The latter remains in its predicted route area to the Jones Falls Bridge, a crossing which it shares with three other roads. Of these three, that in the middle is in a clearly marked secondary route area, an equally direct variant of the primary route that passes five miles northwest of the city. The route used has more

⁸⁸ In some cases, for ease of construction, primary route areas on Map One are left unshaded in their final approach to the shoreline.

Table 1: Summary of Roads Predicted and Not Predicted

Miles of road ¹	Percent of total miles ¹	
688	41.0	Completely predicted. (contained in primary route areas) ²
434	25.9	Partially predicted. (following secondary routes)
213	12.7	Not predicted, but explained by hypothesis when secondary factors are included. (indirect ridge routes, passes, interaction with waterways etc.)
33	1.9	Not included in hypothesis. (turnpikes)
309	18.4	Roads that might contradict the hypothesis of ridge roads.
1678	100.0	Total all roads.
<p>¹ All calculations performed in map centimeters before conversion.</p> <p>² Including 45 miles (2.7 percent) in primary route areas as modified by undrained elevations.</p>		

river crossings, but it uses the same bridges as other routes and combining the roads in this way saves work.

The road to the east along the shore cannot be found by the model. It makes a sharp bend below the forks of a river and then around the head of the next bay which makes it not technically direct. Like North Point Road, this and the roads following other shorelines are limited by estuaries that rapidly get wide. There may be parallel routes inland that cross fewer streams, but the point where a river narrows to where a bridge can reach across often coincides with the limit to which a boat can ascend the stream. Bridge heads thus exert a double attraction on the road. If these points are to be connected, the road can no longer follow a ridge route. It must either bend, or cross additional streams.

The complications of bridge sites are epitomized by the roads to the south of Baltimore that depend on the very long Light Street Bridge, crossing a wide but shallow bay. The model assumes this is impossible and does not find the roads at map letter K.

Two roads from the southwest approach by predicted ridge routes, but the primary route areas go past Baltimore two and four miles west, while the historic roads join together and turn east. In effect these roads continue toward map letter G using the road east of the Jones Falls. Though this route is briefly ninety degrees out of line, the added distance is slight and the miles of road to be built and maintained is very greatly reduced.

The ridge road from the west is predicted to come down to the bay south of Baltimore, but a shift to the north allows a direct continuation by the shore road, while it still reaches water at the Inner Harbor.

The economies of convergence must be called upon, but no addition to the hypothesis is necessary to fully identify the location of Baltimore. It is clear that roads would form a major junction here without the intervention of any other force but the action of terrain on travellers.

Throughout the study area, many of the roads not predicted including those on most of the shorelines, several spurs to harbors, and cases of roads that abandon ridge routes to join with other roads can be explained by the economies of interconnection. The economies to be achieved by sharing facilities gives roads a kind of magnetic attraction for each other that may pull against the natural routes. Once a junction is established by the crossing of natural routes, a road that goes to that junction can form endless direct routes by more than one continuation. The potential for increased traffic may override cost

factors in the location of later roads. Defining ridge routes isolates these effects for closer examination, but to give them a rigorous definition lies far beyond the scope of this paper. Where the economies of convergence are obvious, the roads that respect them should not be taken to contradict the hypothesis.

Other Roads Not Predicted

One road on Griffith's map follows almost a straight line. Frederick Turnpike (map letter D) runs close beside a correctly predicted ridge road, but is below the top of the ridge and crosses a large number of streams near their headwaters. This was the first road in Maryland to be built from scratch as a public project, and was opened the same year that Griffith's map was published.⁸⁹ With a large capital investment and a large traffic, the cost of a bridge declines relative to the cost of detour and the balance of costs has a different solution. Frederick Turnpike is not a primitive road and the hypothesis does not apply to it.

There remain only a few miles of roads that are in no sense predicted by the model and cannot easily be explained by the hypothesis of ridge roads. These roads, the only ones that could contradict the hypothesis, are few. They account for only 18.4 percent of the mileage of early roads in central Maryland.

False Predictions

Routes that are predicted by the model but not occupied by historic roads may also indicate that the hypothesis is not a complete explanation of the cause of location of roads, but these are somewhat more difficult to define and analyze than are roads that are not predicted. Whether an historic road matches a predicted route is easily seen because the historic roads are clearly defined and have a definite mileage. Whether a predicted route matches an historic road is more difficult to decide because the predicted routes are less sharply defined, and their total mileage is not given by the model.

Headwater and confluence curves define areas and not lines. When the boundaries of route areas are close together, the position of the route is defined with precision. When they are far apart, the entire area between them contains equally direct endless routes that

⁸⁹ Gary Lawson Browne, *Baltimore in the Nation; 1789-1861*, (Chapel Hill: Univ. of North Carolina Press, 1980).

all cross the same number of rivers. The boundaries imply that the natural route is in the area, but not that the entire area should be road.

Sometimes a route area has a sort of island in it, where equally good routes run on both sides of an obstacle (map letter G). A choice between the two sides may be based on small scale differences. The unused alternative is not a false prediction, but an imprecision.

Alternate versions of a route also exist where one is a secondary route with an additional bridge. If some condition makes the secondary route the location of a road, there is no reason for a second road to use the primary route. This should be counted as only one difference from the prediction, not two.

Unfortunately, alternate versions of routes occur in all lengths. A route may run almost across the map and yet be identified as a secondary branch only because near one edge it converges on a lower numbered route area. The difference between a distinct route, where a road should be expected, and a trivial variant, where none should, can become problematical.

To define the difference between routes far enough apart to be separate roads and routes so close they duplicate each other is to define the density of the road network. This the model does not do. It distinguishes natural routes from their immediate surroundings, but it does not say how many of these routes should become roads.

The density of routes on the predicted map is limited by selecting the primary routes, those that remain distinct from edge to edge. This definition is convenient, but it is affected by the arbitrary edges of the map. Taafe, Morrill and Gould provide a formula by which road density could be derived from population,⁹⁰ but to use it would add a complexity of dubious relevance. The density of roads might just as well be taken as an independent given and read from Griffith's map. A route area does not predict the location of a road if an historic road runs close beside it. How close can be defined by the normal spacing between historic roads. Unfortunately this approach tautologically disposes of all false predictions. No road could be added to the existing network without increasing its density.

⁹⁰ Taafe, Morrill and Gould, *op. cit.*

Without an independent definition of road density, there is little that can be said about predicted routes not used by roads. A few unused primary routes run cross-country in the west where the historic roads all run radially from the bay. Some are used by later roads. That at map letter C is now State Highway 27. There seem to be few routes that should be expected to be roads according to the hypothesis of ridge roads, and that do not appear as roads on Griffith's map.

Sources of Error

A match between the derived map of natural routes and the reconstructed map of early roads is not exactly the same as a match between natural terrain features and the roads used by travellers. Each of the two maps is the product of a series of several steps with opportunities for error at every step.

Possible Errors in the Map of Ridge Routes

Construction of the contour map of travel cost began with a map of the rivers and coastlines extracted from the modern topographic map (USGS 1:250,000). This map does not describe the river system of 1794, but differences are minor. Maryland is not a delta region where the rivers change rapidly. The largest changes are man-made where urban streams have been covered over, or bays filled in. Using maps of intermediate date,⁹¹ some early details were restored.

There is also a difference between the rivers on the land and the rivers on the map. The river system on the geodetic map is generalized, and the standard is inconsistent. The density of rivers is noticeably less in areas where other categories of information compete for map space. Some rivers could be reconstructed by interpolation from contours, but a consistent level of generalization could not be achieved. Further losses of detail might have occurred in copying the river system onto an overlay.

Errors of this kind can have only minor effects. If a missing river were restored to the map, some secondary route might then become an equal variant of a primary route, but the boundaries of the route area would not change. Errors in the map of rivers can change the

⁹¹ Browne, pp. 6, 108, 150, 188.

order of routes, but it would take many such errors conspiring to change the position and alignment of route areas.

Other errors may arise in the construction of the contour map of travel cost. In particular, when the width of a route area approaches the thickness of the drafted lines, deciding whether the route area is continuous cannot be completely objective. Nevertheless, errors here can have only the same minor effect as errors in the map of rivers.

The arbitrary edges of the map generate errors in two ways. First, a route from edge to edge is not necessarily endless. If a route falls into a pit just beyond the edge of the map, it is not a natural route. The possibility of such errors is minimized, by the way route areas spread out near the edges of the map. Most of the edge is contained in either primary or secondary route areas. These areas narrow as they enter the map, and act like funnels to direct any route from the edge into one or another of the predicted routes.

A second edge effect arises when the edges of the map are not straight and parallel. In this case, there is a risk that the minimum number of river crossings will merely identify the narrowest part of the map. There is only one case in the study area, mentioned on page 78, where this effect can be discerned. The local minimum number of river crossings is defined by comparison only to adjacent areas, and route areas in the narrower part of the region are not adjacent to those in the wider part. Clearly defined routes in the wide part have road value numbers far higher than those on shadow areas in the narrow part.

This second edge effect would be eliminated if all edges were straight and parallel, but this is a geometric impossibility for more than two orientations. Furthermore, straight edges introduce a third kind of error where a river crosses the edge and leaves the map. Routes approaching this point all must cross the river, but some cross it before they leave the map and so have more crossings counted than do those that leave the map before they cross the river. This effect is eliminated if edges are rivers.

To minimize all edge effects, a region should be smoothly rounded with boundaries formed from rivers. The borders of central Maryland do not fulfil these criteria, but where the straight line boundary of Pennsylvania forms part of the edge of our region we have simply continued the predicted map across the line to a river boundary, and then trimmed it off at the state line.

Though there are numerous sources of error in the chain of connection between the natural surface and the predicted map, there is no reason to doubt that the contour map of travel cost is a reasonably accurate representation of where endless, direct routes that cross a minimum number of rivers exist on the land.

Possible Errors in the Map of Early Roads

More serious sources of error lie in the difference between the roads that were on the land in 1794 and their representations on the map that we have reconstructed from Griffith's map. Griffith's record of rivers and coast lines is accurate enough to suggest that his observations of roads are equally reliable. The large scale distortions arising from the lack of trigonometric control do not impugn the accuracy of the underlying observations.

The weakest link in the chain is the transfer of the early map onto the modern base. This reconstruction depends on the assumption that if a modern road has the same general shape and location as an historic road, it is the same road. Though this is not entirely a safe assumption, A. Flora has demonstrated for a section of Delaware that the roads mapped in 1953 are virtually identical to those mapped in 1868.⁹² The assumption that roads are inertial should probably be accepted in the absence of contradictory evidence.

For the most part, there is little difficulty in finding modern roads that corresponds to those on the early map, and these can be identified with confidence. In a few cases, a road on Griffith's map has no modern equivalent. Here the early road can only be approximated from the shape on the early map. Such a reconstruction is therefore less than certain. One short road on Griffith's map has been left off the reconstruction for want of adequate evidence of its location (map letter B). Some subjective judgements were necessary in reconstructing the map and it might seem that these would give latitude for specious interpretation. In fact, there is a strong tendency for the roads most completely predicted to be the same roads most confidently reconstructed.

A thorough study of historical sources would no doubt change the map of early roads, but if the reconstruction were completely inaccurate, it would still demonstrate the existence of a class of roads that follow ridge routes.

⁹² Andrew Flora, correspondence with the author.

Though there are many possible sources of error in the reconstruction of the early map and in the construction of the contour map of travel cost, none of these errors is likely to be large. Furthermore, errors would be far more likely to cause a mismatch between prediction and observation. The potential for error cannot undermine the demonstrated conclusion that in 1794 the roads of central Maryland generally followed ridge routes.

The Cause of Location of Roads Defined

The correlation between ridge routes and early roads is not perfect, but it is clearly significant. Ridge routes are the implications of a physical and economic connection through which terrain could act on the location of roads.⁹³ There is no question that this connection operates, or that it operates in only one direction. The question is whether it acts with sufficient power to override other forces.

Forty one percent of the mileage of early roads in central Maryland are completely explained by the connection between terrain and travel cost. If some other force is acting on these roads, it has either no effect or exactly the same effect. If it can be assumed that a cause with the same effect is the same cause, the fact of a causal connection is proven.

There is room for other forces to modify the locations of some roads, and some of these can be identified. A small number of roads may be largely determined by conditions that we have not considered. Nevertheless, these can only be of secondary importance. As Newton said, "No more causes of natural things are to be admitted than such as are both true and sufficient to explain the phenomena of these things."⁹⁴ Since ridge routes are sufficient to explain most of the facts, other causes are not to be admitted. Most early roads in central Maryland are determined by the shape of the land. Ridge routes cause the location of roads.

⁹³ There remains an ambiguity in that there is no simple way to distinguish between seeking the advantages of the well drained high ground from avoiding the disadvantages of the rough and soft land around the rivers. Nevertheless, making such a distinction would be splitting hairs. It is the whole pattern of difference between ridge and valley that causes early roads to follow ridges.

⁹⁴ Isaac Newton, *Principia*, quoted in "Cause," *Encyclopaedia Britannica*, 1910, V, 558.

Chapter 5

Generalizing the Results

A description of cause and effect in a form that allows prediction is a natural law. If it explains geographic facts it is a geographic law. The hypothesis of ridge roads has been tested and its implications match the facts, but whether the match between early roads and ridge routes is strong enough to prove a law remains arguable. More testing is needed. One important test is a more accurate reconstruction of the map of early roads based on thorough research and field archaeology. Another is to apply the model to more and larger regions. All law is tentative, subject to revision, refinement, and reinterpretation. Confirmation and acceptance are as important to law as elegance. Nevertheless, the hypothesis of ridge roads can now be called the law of ridge roads.

Limiting Assumptions

We have not shown that this law applies to any except the roads in central Maryland in 1794. Nevertheless, this law is general. It takes a large number of disparate facts, the twists and turns of a sizable road system, and makes them one fact, roads follow direct routes that avoid rivers.

It is probable that many other roads follow ridge routes, yet clearly, not all do. Roads in passes and valleys were already known, and we found, in Maryland, turnpikes and shore roads that do not follow ridge routes. Like the law of gravity, this law can be overridden by a sufficient opposing force, and opposing forces are many. A law assumes conditions as part of its definition, and does not apply where these conditions are not met.

The four statements with which we defined the hypothesis of ridge roads contain seven underlying assumptions. The law of ridge roads can apply only when all of them are valid. To define the range of application of this law, it is necessary to examine, in turn, the limits of each of its constituent assumptions.

Rivers and Slight Relief

Geographic laws are commonly regional characteristics. Rivers flow to the sea in Europe, but not in Nevada. That rivers flow to the sea is a sound law, even though the Humboldt, the Chari and the Tarim are in a class to which it does not apply. An assumption about water supply is built into this law. Similarly, the law of ridge roads contains the assumption that resistance can be equated with rivers. This assumption is valid only in humid regions of slight relief. It defines the geographic range of the ridge road.

Humid regions of slight relief can be defined as regions where the shape of the land is dominated by rivers and their valleys. This definition excludes mountain regions where the peaks are the largest terrain features, dry regions without rivers, and, also, true lowlands where the rivers have no valleys and roads follow natural levees.

By rough rule of thumb, all areas below a thousand feet of elevation and above ten with more than ten inches of rainfall are humid regions of slight relief, but what this definition gains in precision it loses in accuracy. Elevation is not the same as relief. Hills with gradients sufficient to affect roads can be found at low elevations. In Connecticut, there are clearly defined passes between "mountains" only a hundred feet high. On the other hand, mild gradients may occur at high elevations. There is a ridge road in Luzon at six thousand feet, and much of Abyssinia is reached by ridge roads at similar altitudes. The range of the ridge road might correlate to some measure of local relief or slope, but the only definition of humid regions of slight relief that is both practical and accurate is the tautological one, that they are regions where ridge roads occur.

The model defined here provides a simple, straightforward method by which the geographic range of the ridge road can be established. Contour maps of travel cost can be constructed for the entire world. They can be compared to the map of roads to discover

which roads follow ridge routes. This task is straightforward, but not small. At my speed, mapping the land surface of the world will take three hundred years around the clock.⁹⁵

Smaller scale maps do not help. A river system must be generalized to map it at a small scale. Ridge roads are often more affected by the small rivers they can avoid than by the longer rivers with no way around. These small rivers disappear from most maps smaller in scale than that used here. It remains, therefore, beyond the scope of this paper to attempt a complete and precise definition of the geographic range of the ridge road. Nevertheless, this question is too important to ignore.

One of the virtues of a model is that it educates perception. Having defined exactly what ridge roads are and do, it becomes easier to recognize them wherever they occur. Without producing an elaborate projected map, we can look for ridge roads on maps of many regions and form an educated opinion of what would be revealed by a rigorous measurement of the global surface.

Many regions of the world contain enough clear examples to suggest that the law of ridge roads applies. It appears that ridge roads dominate the American coastal plain from Long Island to central Florida, and the Gulf coast to central Texas. On the lake plain west from Cleveland, they are interspersed among the section line roads. Around the hundredth meridian, valley roads take over. Northwards, ridge roads reach the tree line in Canada.

In South America, ridge roads cover most of the area between the coastal rampart of Brazil and the foothills of the Andes. The road from São Paulo toward Rondônia, for one, is on ridges most of the way.

In Europe, ridge roads cover most of the northern plains that extend east from England to reach the Black Sea and the Altai, interrupted by elevations like the Harz and the Urals. The world's longest perfect ridge route may be the road from Odessa toward L'vov. Ridge roads are common in the Indo-Gangetic plain and also in the Deccan, where the road from Hyderabad to Poona is a ridge road. They also occur on the North China Plain, in the Red Basin, and at high altitudes between Guiyang and Kunming. Throughout most of Africa south of the Sahara, except in the mountains, they are common. In Australia,

⁹⁵ At the present rate of production, a contour map of travel cost for one standard orientation at a scale of 1:250,000 takes about thirty seconds per square mile, after preparation. The model is an algorithm that can probably be computerized, improving the speed, though to do so is not a simple problem.

they occur both inland and on the coast, where the Pacific Highway north from Sydney is a perfect example.

These broadly outlined regions also have many roads that are not ridge roads. Like Maryland they contain isolated hills and swamps, and sub-regions that are not humid regions of slight relief. Everywhere there are some roads that do not follow natural routes.

Only occasionally are there clear edges to the regions where ridge roads occur. In Virginia, Maryland, and Pennsylvania, there is a sharp line where the mountains begin. Where the region of slight relief adjoins the region of high relief, ridge roads connect to roads that follow passes and valleys.

Elsewhere there are regions of moderate relief where ridge roads and valley roads are mixed together. In southwest Wisconsin, U.S. Route 18 is on Military Ridge, while the next parallel road follows the Wisconsin River Valley. These two roads are connected by a number of perpendiculars, some in valleys and some on ridges. Around Pittsburgh, ridge and valley roads form an almost regular alternation. The boundaries between regions of high, slight, and moderate relief can be extremely subtle. The difference is often between individual routes rather than between regions.

Overall it appears that humid regions of slight relief, where rivers are the principal resistance to travel and roads follow ridges, include about half the land area of the world and a much larger proportion of its total population and road mileage. It also appears that in most of these regions, not all roads, but most roads, follow ridge routes. The law that the ridge points the road appears to describe the majority of all roads.

Natural Routes and Primitive Roads

Primitive roads follow natural routes, but other kinds of roads may not. The assumption that the capital and operating costs of a road can be described by overland travel cost applies consistently only to primitive roads made of nothing but the natural surface of the land.

It is appropriate to identify the primitive road as a distinct mode of transportation. The difference between modes is more conventionally conceived as a difference between vehicles, for example, the wagon and the boat. The essential difference between modes, however, lies in the form of the physical connection between the traveller and the earth.

The wagon rolls and the boat floats. Because it is through this connection that the surface acts on travellers, changes in mode change the way the shape of land acts on the travellers' choice of route.

Different modes follow different kinds of routes. This is obvious in the case of the wagon and the boat, but there are similar, though subtler, differences between the different kinds of roads, for instance, between the wagon road and the railroad. The rolling surface of a railroad is not the land itself, but a structure built on top of the land. Thus the physical connection between the traveller and the land takes a different form. This difference leads the wagon road and the railroad to follow different kinds of routes.⁹⁶

A modern highway may also be a structure built on top of the land. Primitive roads and highways follow distinctly different kinds of routes, even though they may be used by the same vehicles, and even though one may turn into the other by gradual stages. Eventually roads may be built where overland travel is impossible, like the Mont Blanc Tunnel under the highest point of the Alps, or Lake Shore Drive half a mile out in Lake Michigan.

Though roads with a substantial built structure are not confined to natural routes, they still find them useful. Even with a high capital investment, it tends to remain true that a road will be cheap to build and operate if it is located where the land is firm, smooth, and level in its natural condition. The Mont Blanc Tunnel and the Pulaski Skyway exist only

⁹⁶ James E. Vance Jr. ("The Oregon Trail and Union Pacific Railroad: A Contrast in Purpose," *Annals Assn. of Amer. Geographers* LI, 357-379 [Dec. 1961]) has described in detail one example of how different modes of transportation follow different kinds of routes, specifically, the Union Pacific Railroad and the Oregon Trail across Wyoming. In this dry region the wagon road treats the desert as resistance and the railroad does not. Mountains are a secondary factor. The Oregon Trail makes a long detour to the north where forage and surface water are available. The railroad follows a more direct route farther south across the heart of the desert. Vance identifies this difference with a difference in the purposes of travel, calling the railroad "economic" transportation and the emigrant trail "non-economic," but this explanation cannot be correct. This part of the Oregon Trail was originally opened by pack trains and wagons of the definitely economic, and even fiercely competitive fur trade (Bernard DeVoto, *Across the Wide Missouri*, [Boston: Houghton Mifflin, 1947]), while the railroad too was used for both commerce and emigration (Robert Louis Stevenson, *Across the Plains*, [New York: Scribner's, 1900]).

The real difference that Vance describes lies in the form of the physical connection between the traveller and the land. In particular, the two modes differ in the means by which energy from the environment is converted into mechanical force. In one case the connection is from range grass, to ox, to hoof, in the other from coal mine, to firebox, to driving wheel. The far greater capacity of the railroad to store and transport energy makes it largely independent of its immediate environment for this need. While travellers on the wagon road "had to be more concerned with feed for their animals than with any other aspect of the natural environment," (Vance p. 367) the railroad was free to respond primarily to distance and gradient. Vance offers no evidence that "the location is determined by the use made of the route," (Vance p. 379) but he gives a clear illustration of how different modes of transportation respond to different forms of resistance.

because natural routes are not available nearby. Natural routes are advantageous for almost any kind of road, but only primitive roads follow them with predictable consistency.

There are some routes determined by terrain that are not the natural routes of primitive roads. The Mont Blanc Tunnel is located at the narrowest point of the Alps. Its location is precisely determined by the closest approach of the valleys on opposite sides of the range. The Hoosac and Lötschberg tunnels are similarly sited. If all routes determined by terrain are to be considered natural routes, then the term must be carefully qualified. The natural route of a base level tunnel is not where overland travel cost is least. The term natural route is more precise, and probably more useful, if it is reserved for the natural route of a primitive road. In this narrow sense, natural routes are where overland travel is easy.

Land Cost and Early Roads

The assumption that overland travel cost describes all costs of a road implies that the cost of clearing the land is insignificant. In an empty country, only natural vegetation need be displaced and this cost does not much vary between one location and another.⁹⁷ The situation is different in a region that is already occupied. Clearing the land may mean expensive land purchases and the demolition of entire city districts. Here, land acquisition may become the most important cost of all. One effect of land price can be seen in a circle of railroad terminals at the edge of mid-nineteenth century urban development. This pattern occurs in nearly every major city, but Paris offers the purest example.

Different kinds of roads are differently affected by land cost. Deep subways may be virtually immune. Expressways in cities are all but completely determined by the location of low-priced land. They cut through parks, abandoned ports, and rail yards, and smash through minority communities. Or they follow isolines of development density in a belt around the city. Sometimes they follow a kind of anti-natural route on marshlands or hillsides, land that lies vacant precisely because it is difficult to build on.

The effect of conflicting land use also shows in survey line roads. The section line roads of the western land survey are laid out to avoid prospective occupation and the

⁹⁷ Prairies make better roads than forests, but this choice is seldom available within a local area.

regular grids of city streets are the same thing at a smaller scale. Such roads appear at an early date in eastern Pennsylvania but not in Maryland. Even these roads are not independent of terrain. The roads of Iowa, for instance, seem to select just those section lines that coincide with ridge routes.

Before the Northwest Ordinance, roads in thinly settled country could be located without reference to other uses of the land. Since roads in some form are essential to human use of the land, there is, in every region, a first network of early roads laid out before there is any conflicting land use. Later, as the land fills up, the easiest way to find a clear route is to follow the one that has been in use since the first settlement.

Occasionally a road may be relocated, or even disappear, but once a road becomes hemmed in by development it becomes highly inertial. The cost of clearing the land tends to cause early roads to stay in place and be continually improved. Later roads and later development accept them as given. Early roads form the basic structure of the modern road network.

The assumption that haulage and construction costs are closely dependent on overland travel cost applies primarily to primitive roads and this includes most early roads. The assumption that these costs are the only costs implies that the cost of clearing the land is insignificant and is limited to roads laid out in empty country. This too includes most early roads. Natural routes may be used by modern roads, but only early roads use them consistently.

General Assumptions

Two assumptions contained in the law of ridge roads restrict its application to early roads and a third to humid regions of slight relief. The remaining assumptions are more general in their nature. These are that roads follow economic routes, that costs are described by distance times resistance, and that neither variations in traffic potential nor the interactions of different roads can have an effect on road location. These are not limited to any narrow class of road or region, yet all have other limitations.

Economic Routes

That all roads follow economic routes, there is no reason to doubt. The assumption that roads are located where they create the most benefit at the least cost is equivalent to the

assumption that roads can be described by law. Road builders do not work with perfect information nor solve economic problems with infallible precision, and uncertainties arise from these sources, but these are as nothing compared to the uncertainties of second-guessing their decisions from afar.

Though economic rationality is a cornerstone of human geography, it is dangerous to take too narrow a view of economic values. Many factors are economically connected to the location of roads. Which of these carry significant weight changes with changes in economic, geographic, and technological conditions. Their relative importance cannot be identified by a priori reasoning but only by observing the facts. Though all roads follow economic routes, economic routes are natural routes only when the cost of travel on the natural surface dominates the economic balance. As the economy of transportation becomes complex, it is subject to subtle variations. For instance, when capital institutions are created to build roads, the costs and benefits determining the route are those of the institution. The geopolitical strategies of American rail barons or the elegant rationality of French bureaucrats may not match the interests of travellers.

Though it cannot always be safely assumed that the character of the natural surface is important, or even that roads are located where the most travellers reach their goals with the least work, the physical work of moving a load never becomes irrelevant. That which does not fulfill the basic needs of moving travellers is not a road.

The Endless Road

The statement that a road must be endless is not itself an assumption, but a conclusion drawn from two assumptions. The first is that variations in traffic potential have no effect on the location of roads. This assumption leads directly to the inference that traffic is proportional to area. The conclusion that roads must be endless follows only with a second assumption that interconnection has no effect and that benefit can be described for a single route in isolation. It is only this last assumption that leads to the most common exceptions to the rule that roads are endless.

When natural routes surround some area on all sides, any long distance traveller has an easy way to bypass the area. No through route in the area can attract these travellers, and added extension will not increase traffic. The area still generates traffic, but with the local volume small, fixed costs outweigh operating costs and the economic route is the shortest

distance to a passing road. This solution, of course, is normally perpendicular to the through route. The fringe of driveways, farm lanes, and dead-end streets that line the sides of highways do not necessarily imply the presence of some attraction at the end. We have called them the distinct phenomenon of lanes, but they are in fact a distinct solution to the same area per road equation that implies endless routes. The assumption that variations in traffic potential have no effect on the location of roads is sweeping, and there are many reasons to doubt that it is universally true. It may be no more than a convenient abstraction to simplify the problem. But in some sense it is broadly true of roads in general, so long as location is kept distinct from density. Roads serve areas, not points.

When roads are endless, they are described by endless routes, and the common conception of a road as a link between endpoints becomes seriously misleading. It confuses the road, which has no endpoints, with the trip, which does. If applied to the cause of location, this definition implies the false assumption that cities and junctions are independent conditions. Even when applied to existing networks, it interferes with understanding. If a road network is taken apart at junctions, it is not analyzed according to a consistent organizing principle, but merely chopped up.

The endless route with a near constant orientation is an alternative conception of the nature of a road. Endless extension describes with extreme simplicity the conditions under which a road serves the greatest traffic. It describes the essential difference between major and minor routes. The Clark Fork Valley, for instance, though a narrow and difficult route, is yet important because it connects the Columbia to the Yellowstone to form the through route followed by Lewis and Clark, two railroads, and a U.S. and an interstate highway. The adjoining Bitterroot Valley, wider and more populous, is an insignificant route because it leads essentially nowhere.

The through route across the map is the logical intersection of an endless route and a region. It analyzes a road network into component parts that can be separately examined without destroying their continuity. Classifying roads by orientation fits all roads into a small number of significantly differentiated categories. It analyzes not only existing networks, but all possible locations for roads. The endless route is a logical conception of the nature of a road adequate to the problem of analyzing the cause of location of roads. If exception is made for lanes, endless extension is the essential nature of a road in the broadest sense of the word.

The Limit of Detour

When roads are endless, the trade-off between distance and resistance becomes the central issue in their location. Endless extension defines the benefit side of the economic equation with a constant. The economic variables are all on the cost side, and the dynamics of cost determine the choice of route. Endless roads go where travel is easy.

Distance times resistance describes in a general way, not just overland travel cost, but nearly all the costs of transportation from engine wear to snow removal. As distance rises, the costs of haulage, construction, land acquisition, and maintenance all rise with it. To minimize distance roads are direct, but a straight line will often strike resistance that can be avoided by a trivial detour. When resistance varies with location, roads find a balance between distance and resistance.

Resistance takes many forms. Different environments present different conditions that make travel difficult. In mountain areas, roads avoid gradient. We have shown how some roads avoid rivers. Others avoid deserts or swamps or conflicting land uses. Open water is a resistance to land roads almost too obvious to notice. All these are forms of resistance to roads.

What conditions constitute resistance also varies with the mode of transportation. Surface material has a strong effect on primitive roads but very little effect on railroads, which are far more susceptible to gradient. Expressways, because of their great width, are extremely sensitive to the price of land. Subways are much less so. The Europe to Japan air routes that, until recently, ran via Anchorage acknowledged as resistance the political frontier of the Soviet Union. Ships crossing the North Atlantic, like the Titanic, seek a balance between the minimum distance of the great circle and the resistance of the southern limit of floating ice. Roads follow natural routes when the resistance to which they respond is that presented by the natural surface of the land. When they respond to other forms of resistance, they follow other kinds of routes.

Whatever resistance roads avoid, there is always a limit of detour. If resistance is finite, there is a limit to the distance worth travelling to avoid it. This limit is by no means a constant, but varies with terrain, economic conditions, and kind of road. Mountain roads detour more than those in plains; modern roads detour less than early roads. The limit of detour is not quite constant even within the relatively homogeneous road network

of federal Maryland. Every combination of mode and environment has a different solution to the trade-off between distance and resistance.

The limit of detour can be defined and measured by an angle only if the road is nearly straight. Other forms and relative degrees of resistance could be introduced and the measure used in the model could be varied, but as the angle approaches ninety degrees its projection into a boundary ceases to differentiate places. Changes of local orientation much larger than this are common in switchbacks or in roads that follow rivers. Generally, whenever resistance is localized and the rest of the land is passable, roads are direct enough for an angle of detour to describe them. This model cannot be used where passable land is the exception, but there it is not needed.

All roads follow direct routes that avoid resistance. Though the meanings of “direct” and “resistance” vary widely, the trade-off between distance and resistance describes all transportation costs and is a universal law of road location.

Of the four statements with which we defined the hypothesis of ridge roads, three are true of virtually all roads. Roads are located to create the most benefit at the least cost, they create the most benefit when they are endless, and their cost can be described by distance times resistance. These assumptions are not universally true – there are important exceptions – but of all persistent paths of travel, all roads in the widest sense of the word, most follow endless, direct routes that avoid resistance. When the resistance comes from nature, roads follow natural routes.

Natural Routes and Geographic Law

When the ridge routes of humid regions of slight relief are added to the passes and valleys of high relief, the natural levees of very low relief, the kind of one-sided valley of the mountains’ base, the strings of desert waterholes, and the navigable waterways, natural routes cover the earth with a fine network. Every part of the earth contains locations that are determined by the shape of the land to be the best places for travellers to pass through.

The fact that natural routes are permanent features of the land is not enough to make roads unchangeable in their locations. Changes in economic and technological conditions can shift the advantage between waterways and roads or otherwise lead travellers to prefer one natural route to another. Such adjustments may make an accessible location in one age isolated in another. Much of the material of history, the rise and fall of cities and, with

them, of the kingdoms they rule, may be understood as the effects of these shifts in the patterns of communication. Vidal de la Blache said "*Routes et villes, villes et routes; C'est l'histoire,*"⁹⁸ but if natural routes are as significant as appears, this epigram may be emended to read "Cities and history are natural routes."

When roads follow natural routes, all that follows roads follows the shape of the land. The transport conditions of human history define the distributions of races, languages, and cultures. Natural routes link physical and human geography. They describe a force acting from below, from plate tectonics through geomorphology, shaping the space within which people act and focusing their choices with a determining power equivalent to that by which the sun acts through climate to determine natural vegetation and narrow the choices of agriculture. The geographical determinists of the turn of the century underrated the significance of natural routes, finding them only in mountain areas. Including them with climate among the influences of environment may still not explain all the facts of human geography, but it moves the field closer to explaining "the manner in which the natural phenomena, including man, [are] distributed in space."⁹⁹

Morphological Law

"To explain [these] phenomena . . . means always to recognize them as instances of laws."¹⁰⁰ Geographic law describes the cause of location. Since the cause of location must be a difference between locations, geographic law describes functional relations between geographic distributions. Geographic law is morphological. It explains the shapes that appear on the map.

Such law is not quantitative. It is doubtful that shape can be quantified, but quantification is not essential to describing cause and effect. The law of ridge roads defines causal conditions with qualitative distinctions that can be read from the map. Through a logical rule that transforms one map into another, it describes how they act. The few quantities that appear, the measure of the limit of detour, the count of river

⁹⁸ Vidal de la Blache, quoted in Fernand Braudel, *The Mediterranean and the Mediterranean World in the Age of Philip II*, trans. Sian Reynolds, (New York: Harper, 1972), I, 76.

⁹⁹ Schaefer, p. 228.

¹⁰⁰ Schaefer, p. 227.

crossings, and the total mileage of roads, are peripheral and dispensable to the operation of the model. The prediction is expressed as a map.

The basic tools of cartographic logic have long been known. There is nothing new about Köppen and Thünen. Applying their methods to the problem of road location is largely a matter of detail. The map is an analytical tool that provides mathematical rigor without calculation and describes shape without quantification. It yields precise results in terms of location, the terms in which geographic questions must be answered. Though the ultimate potential of cartographic logic remains to be seen, the logic of the map provides a calculus of shape through which to define and test morphological law.¹⁰¹

The map is the essential and irreplaceable model of spatial relationships. Schaefer says, "One hardly needs to elaborate how useful it is to have such a convenient and selfregulating [sic] vehicle for the process of abstraction."¹⁰² And yet the map is not at the core of modern geography. One recent review of geographic methodology mentions maps only in a passing reference to "the reasons for the relative decline in the use of maps in the discipline."¹⁰³

Appleton says:

The morphological approach has flagged for want of general theory and good measuring devices. In transport geography, too, some advances must be made in these directions if worthwhile results are to be achieved.¹⁰⁴

Though it is true that the morphological approach has flagged, it is not for these reasons. General theory comes from describing basic cartographic techniques in terms of set theory and the Venn diagram. The superimposition of maps is a measuring device. Like the pan balance scale, it measures the unknown by comparing it to the known. The scale measures weight against weight: the map measures shape against shape. If the

¹⁰¹ The contour map of travel cost may prove to be applicable to the design of transportation systems. Identifying routes by logical elimination from the surface of the map performs by a rigorous method a step in the design process that has always been performed by intuition, by trial and error, or by the mere repetition of past solutions.

¹⁰² Schaefer, p. 244.

¹⁰³ James Bird, *The Changing Worlds of Geography: A Critical Guide to Concepts and Methods*, (Oxford: Clarendon Press, 1989), p. 34.

¹⁰⁴ J. H. Appleton, *A Morphological Approach to the Geography of Transport*, (Hull: Univ. of Hull, 1967), p. 13.

morphological approach has flagged, it is more likely because shape is not quantitative and so cannot satisfy shallow preconceptions about the science of cause and effect.

A Single Fact

Temerarious extrapolation aside, only one fact is here proven. Some roads follow ridge routes. When endless direct routes that avoid rivers are mapped in central Maryland, they match the early roads with a significant level of consistency. Because there is a connection through which the land could act on roads, and because the shape of the land is independent of roads, cause and effect can be inferred.

Terrain is not the only force that acts on roads. Ridge routes cannot completely explain the facts even in this one region. Of the forces that oppose them, some can be identified, others not even guessed. Finding ridge routes isolates one major force, and makes the forces that remain so much easier to identify and define. The significance of these other forces is not to be disparaged, but the common idea that roads are determined by terrain only in mountain areas is clearly a misconception.

Central Maryland is not a mountain area. It is a region of slight relief of the kind of which it has been said that roads "run in any direction expediency suggests."¹⁰⁵ Nevertheless, roads here follow ridges. Ridges that form direct routes are specific features of terrain that are the best places for travellers to pass through. They are natural routes. Natural routes exist outside mountain areas.

¹⁰⁵ Semple, p. 548.

References Cited

- Abler, Ronald, Adams, John S., and Gould, Peter. *Spatial Organization; The Geographer's View of the World*. Englewood Cliffs: Prentice Hall, 1971.
- Appleton, J. H. *A Morphological Approach to the Geography of Transport*. Hull: Univ. of Hull, 1965.
- Beckmann, Martin. "Principles of Optimum Location for Transportation Networks." in *Quantitative Geography, Part 1 Economic and Cultural Topics*. Evanston: Northwestern Univ. Dept. of Geography, 1967.
- Bierce, Ambrose. *The Enlarged Devil's Dictionary*. Ed. Ernest J. Hopkins. New York: Doubleday, 1967.
- Bird, James. *The Changing Worlds of Geography: A Critical Guide to Concepts and Methods*. Oxford: Clarendon Press, 1989.
- Bird, J. E. *Centrality and Cities*. London: Routledge Direct Editions, Routledge Kegan and Paul, 1977.
- Braudel, Fernand. *The Mediterranean and the Mediterranean World in the Age of Philip II*. Trans. Sian Reynolds. New York: Harper, 1972.
- Brown, Lloyd A. *The Story of Maps*. Boston: Little Brown and Co., 1949.
- Browne, Gary Lawson. *Baltimore in the Nation; 1789-1861*. Chapel Hill: Univ. of North Carolina Press, 1980.
- Brunhes, Jean. *Human Geography: An Attempt at a Positive Classification, Principles and Examples*. Trans. I. C. Lecompte. Chicago: Rand McNally, 1920.
- Bruton, Michael J. *Introduction to Transportation Planning*. London: Hutchinson, 1985.

Christaller, Walter. *Central Places in Southern Germany*, Trans. Carlisle W. Baskin. Englewood Cliffs: Prentice Hall, 1966.

Cooley, C. H. "The Theory of Transportation." *American Economic Assoc.* IX, 3 May 1894, in Eliot Hurst.

Cooper, James Fenimore. *The Last of the Mohicans*. 1826; Philadelphia: Porter and Coates, no date.

Cornish, Vaughan. *The Great Capitals: an Historical Geography*. London: Methuen, 1923.

DeVoto, Bernard. *Across the Wide Missouri*. Boston: Houghton Mifflin, 1947.

Dunbar, Seymour. *The History of Travel in America*. 1915; reprint New York: Greenwood Press, 1968.

Eisenberg, Murray. *Axiomatic Theory of Sets and Classes*. New York: Holt, Rhinehart and Winston, 1971.

Eliot Hurst, Michael E., ed. *Transportation Geography; Comments and Readings*. New York: McGraw Hill, 1974.

Encyclopaedia Britannica, 1910-11, V, XXIII.

Garrison, W. L., and Marble, D. F. "Graph Theoretic Concepts." reprinted from *A Prolegomenon to the Forecasting of Transportation Development*, Research Reports, Transportation Center, Northwestern Univ. 1956 pp 40-68 and 89-96, in Eliot Hurst.

Goodall, Brian. *Penguin Dictionary of Human Geography*. Harmondsworth: Penguin, 1987.

Griffith, Dennis. "Map of the State of Maryland." 1794 [1795]. in Edward C. Papenfuse and Joseph M. Coale III, eds., *The Hammond - Harwood House Atlas of Historical Maps of Maryland, 1608 - 1908*. Baltimore: Johns Hopkins Univ. Press, 1982.

Hartshorne, Richard. "The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past." *Annals Assn. of Amer. Geographers*, XXIX, 3 Sept. 1939.

Herodotus. *The Histories*. Trans. Aubrey de Selincourt. 1954; Harmondsworth: Penguin, 1972.

Ibn Hauqal. *Configuration de la Terre (Kitab Surat al-Ard)*. Introduction et Traduction par J. H. Kramers et G. Wiet. Paris: Editions G. P. Maisonneuve et Larose, 1964.

Kansky, Karol. *Structure of Transport Networks, Relationships Between Network Geometry and Regional Characteristics*, Dept. of Geography Research Paper No. 84. Chicago: Univ of Chicago, 1963.

Kissling, C. C. "Linkage Importance in a Regional Highway Network." *Canadian Geographer* XVIII, 2 1962 pp. 113-129. in Eliot Hurst.

- Kolars, J. F. and Malin, H. J. "Population and Accessibility: An analysis of Turkish Railroads." *Geographical Review* Vol. 60 1970. pp. 229-246, in Eliot Hurst.
- Lewis, Pierce. "New Orleans, The Making of an Urban Landscape." in Adams, John S., Ed. *Contemporary Metropolitan America*. Cambridge: Ballinger Publishing Co., 1976.
- Mitchell, Jean B. *Historical Geography*. London: The English Universities Press, 1954.
- Patton, Clyde P., Alexander, Charles S., and Kramer, Fritz L. "Chapter 19." *Physical Geography*. Belmont, Cal.: Wadsworth, 1970.
- Platt, Robert S. "Central American Railways and the Pan-American Route." *Annals Assn. of Amer. Geographers*, XVI no. 3, (Mar. 1926).
- Schaefer, Fred K. "Exceptionalism in Geography: A Methodological Examination." *Annals Assn. of Amer. Geographers* (Sept. 1953)
- Scull and Heap. "Map of Philadelphia." Engraved by William Faden. 1777; Reproduced Ithaca: Historic Urban Plans, 1969.
- Semple, Ellen Churchill. *Influences of Geographic Environment; On the basis of Ratzel's System of Anthropogeography*. New York: Holt, 1911.
- Smailes, A. E. *The Geography of Towns*. London: Hutchinson, 1953.
- Stanley, W. R. "Evaluating Construction Priorities of Farm to Market Roads in Developing Countries: A Case Study." *Journal of Developing Areas* V, no. 3 (1971), pp. 371-400, in Eliot Hurst.
- Stevenson, Robert Louis. *Across the Plains*. New York: Scribner's, 1900.
- Taaffe, E. J., Morrill, R. L., and Gould, P. R. "Transport Expansion in Underdeveloped Countries: A Comparative Analysis." *Geographical Review*, vol. 53 (1963) pp. 503-529, in Eliot Hurst.
- von Thünen, J. H. *Der Isolierte Staat in Beziehung auf Landwirthschaft und Nationaloekonomie*. 1826; Berlin: Wiegandt, Hempel & Parey, 1875.
- Ullman, Edward L. *Geography as Spatial Interaction*, Ed. Ronald R. Boyce. Seattle: Univ. of Washington Press, 1980.
- Vance, James E. "The Oregon Trail and the Union Pacific Railroad: A Contrast in Purpose." *Annals Assn. of Amer. Geographers*, LI, 4. (Dec. 1961)
- Wellington, Arthur Mellen. *The Economic Theory of the Location of Railways; An Analysis of the Conditions Controlling the Laying out of Railways to Effect the Most Judicious Expenditure of Capital*. 1887; New York: John Wiley and Sons, 1899.
- Wood, J. S. "Elaboration of a Settlement system: The New England Village in the Federal Period." *Jour. of Hist. Geography*, CIV (1984) p. 334.

Appendix A

Glossary

access range : the limit of distance within which a point is near a road and beyond which it is not (p. 26)

alignment : the map shape of a route or a path (p. 3)

angle of detour : the angle between an isosceles path and the straight line of the trip; the base angle of the triangle (p. 43)

base cost (of a trip) : trip length times normal resistance (p. 40)

capital cost (of a road) : the work consumed in modifying the surface of the land to make it easy to travel on (p. 14)

confluence curve : lines projected from the junction of two rivers at thirty degrees to the standard orientation and enclosing the double stream; a boundary similar to the headwater curve (p. 64)

cost : see : base cost; capital cost; crossing cost; detour cost; excess cost; haulage cost; overland travel cost; total cost

cross-country path : a path that does not coincide with a road in any part – it coincides with the least cost overland path between the same endpoints (p. 25)

crossing cost : that part of the excess cost of a path that is attributable to river crossings; the difference between the total cost of a path that crosses rivers and the minimum cost for a path of the same length (p. 40)

defined set of trips : 1 : (set 1) all pairs of points that share a common orientation, trip length, and y co-ordinate of origin (p. 56) 2 : (set 2) set one with a different trip length

3 : (set 3) all pairs of points that share a common orientation and y co-ordinate of origin (p. 56) 4 : (set 4) all pairs of points that share a common orientation (p. 57)

density (of road network) : average distance between roads; mileage of roads per unit of area (p. 3)

destination : 1 : one endpoint of a path or of a trip (p. 9) 2 : a place many travellers go; a geographic condition that acts through the benefit side of the economic equation (p. 19)

detour : the difference between straight line distance and distance following the shape of the route (p. 18)

detour cost : that part of the cost of a trip that is attributable to detour in the path; the difference between the minimum cost for a path of given length and the base cost of a trip; the difference between path length and trip length multiplied by normal resistance (p. 40)

detour ratio : the length of a path divided by trip length (p. 44)

direct : nearly straight (p. 18)

direct path : an isosceles path with an angle of detour less than the limit of detour (p. 46) 2 : a path with a detour ratio less than the secant of the limit of detour (p. 48) 3 : a path that remains within the rhombus of direct paths constructed on its endpoints (p. 48)

direct route : a route of which all segments are direct paths; a route that maintains a near constant orientation (p. 51)

early road : a primitive road laid out on land that is largely unoccupied (p.2)

easy trip : 1 : a trip that has some path on which the cost of travel is as low as on any path for any trip of equal length; a trip for which some direct path crosses no river (p. 50) 2 : a trip with a straight path that does not cross a river (pp. 55-57 only)

economic route : a route on which a road creates the greatest benefit at the least cost (p. 15)

excess cost (on a path) : the difference between the total cost and the base cost; detour cost plus crossing cost; total cost minus base cost (p. 40)

gradient : rise per distance (p. 13)

haulage cost : the physical work of moving a load (p. 13)

headwater curve : an angle constructed with its vertex on the headwater of a river with each leg at thirty degrees to the standard orientation, enclosing the river like a broad

arrowhead pointing upstream – all easy trips past the river have their origins outside the headwater curve (p. 58)

headwater line : a line projected in the standard orientation through the headwater of the river (p. 56)

isosceles path : a continuous curve formed from two equal straight line segments; a path that forms two sides of an isosceles triangle, with the straight line of the trip forming the base (p. 43)

lane : a road that terminates (p. 25)

least cost overland path (for a trip) : the path on which the cost of overland travel is lower than by any other path between a given pair of points (p. 39)

least cost path (for a trip) : the path with the lowest cost of travel under conditions that may include the presence of a road (p. 25)

limit of detour : the angle of detour at which detour cost equals crossing cost; an angle of thirty degrees (p. 46)

line of destinations : the set of destinations of the defined set of trips; a line parallel to and one trip length away from the line of origins (p. 56)

line of origins : the set of origins of the defined set of trips; the line with the given y coordinate and all possible values of x (p. 56)

location (of a road) : 1 : position in space 2 : alignment 3 : relations to other roads; network structure and density (p. 2) 4 : the act of choosing a particular strip of land on which to travel and to build a road (p. 3)

natural path : a segment of a natural route (p. 39)

natural route : a feature of terrain that forms the best place for travellers to pass through; a place where travel is easy; an endless route with the least possible overland travel cost (p. 6)

natural route function : the economic equation of where roads create the most benefit at the least cost – its domain is the surface of the earth, or the map and its range is the ratio of benefits to cost – it describes the historic process of road location (p. 34)

natural trip : a pair of points contained in a natural route (p. 49)

network structure : the relations between connecting roads (p. 3)

normal resistance : the resistance of land away from rivers – resistance cannot be lower than normal resistance (p. 38)

normal ridge route : a ridge route that is not a perfect ridge route (p. 60)

orientation : 1 : the compass bearing between the origin and destination of a trip or path
2 : the general trend of a road or route – one orientation represents two opposite directions

origin : one endpoint of a path or of a trip (p. 9)

overland travel : travel on the natural surface of the land

overland travel cost (of path) : the work required to travel on the natural surface of the land (p. 15)

path : a continuous curve with two endpoints and a specific location at every point; the sequence of locations of a moving traveller (p. 9)

perfect ridge route : a through route that crosses no rivers (p. 60)

primary route area : area with road value number lower than in any contiguous area (p. 64)

primitive road : a strip of land, cleared of vegetation and stationary human uses, but otherwise little changed from its natural condition; a road on which the rolling surface is the natural surface of the land (p. 15)

resistance : 1 : all the conditions that make travel and road building difficult 2 : friction and gradient (p. 17)

rhombus of direct paths : a figure with four equal sides (a rhombus) with angles that measure twice the limit of detour (60 degrees) on the endpoints of the trip – all direct isosceles paths for the trip between these two points are contained within this figure (p. 48)

ridge route : an endless, direct route that avoids rivers; a natural route in humid regions of slight relief; a through route that crosses the least possible number of rivers (p. 62)

road : 1 : any persistent path of travel including airways and seaways; a place that travellers use when making trips 2 : a persistent path of travel on the surface of solid land; a strip of land used for travel 3 : a solid surface used by wheeled vehicles that steer by the friction of their wheels (i.e. not a railroad) 4 : a primitive road (p. 2)

road value number : the number that identifies the least number of river crossings in any through route that enters a given area of the map; an inverse scale of the range of the natural route function (p. 63)

route : the line that describes a road; a continuous curve with a specific location at every point and no endpoints – a route curve (p. 3)

route area : the area that contains a route (p. 10)

route curve : a route (p. 10)

secondary route area : a route area with road value number near the local minimum describing a distinct alternate route in the form of a branch, loop, or crossover between primary routes (p. 65)

standard orientation : the orientation assumed by the defined set of trips (p. 55)

through route : a direct route that crosses the map in the standard orientation (p. 59)

total cost : overland travel cost; base cost plus detour cost plus crossing cost (p. 40)

traffic : 1 : the number of travellers who use a road; a measure of the benefit created by a road (p. 23) 2 : trip miles per road mile (p. 29)

traffic potential : the differences between places that make some the origins and destinations of more trips (p. 24)

travel : the physical act of changing the location of an object; a synonym for transportation

trip : 1 : the act of a traveller changing his location (p. 25) 2 : a pair of points between which a traveller is presumed to move (p. 26) 3 : a pair of points (p. 26)

trip length : the straight line distance between an origin and a destination (p. 40)

Appendix B

Constructing the Contour Map of Travel Cost

The description of the contour map of travel cost in the main text provides no practical way to construct it. Headwater and confluence curves constructed on all the rivers of a large system divide the map into minuscule units that are practically impossible to classify. The following simplified method of constructing this map is logically equivalent, and practical to use.

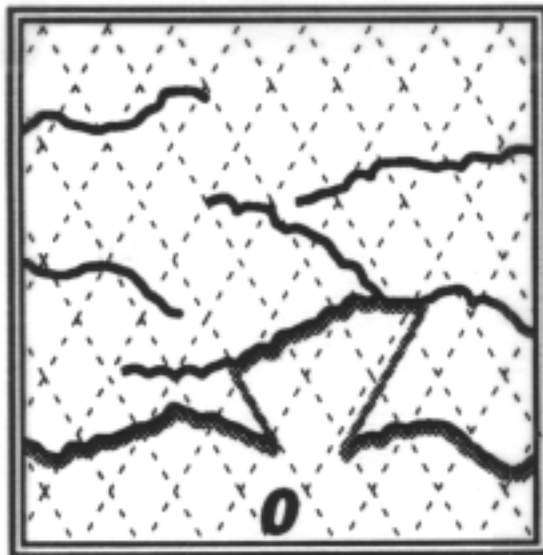


figure 37

The first step is to construct a transparent overlay with two sets of parallel lines, sixty degrees apart. This grid can be superimposed on a map of rivers. When it is squared up so its lines are thirty degrees from the standard orientation on either side, the angles of headwater curves (see page 58) can be traced from this grid. An eighth inch line spacing is convenient.

Next, choose an edge of the map, approximately perpendicular to the standard orientation, to be the starting line. The area near the starting line is a zero-bridge area, any part of which can be reached on a direct route from the starting line without crossing any rivers. This zero-bridge area ends at the first river. Using a first color, construct a boundary for this area by tracing along the river nearest the edge (figure 37). At the headwater of the river, turn back around the river and follow a grid line to form a headwater curve. Extend this line to the next river.

Wherever there is a gap between the rivers nearest the starting line, the zero-bridge area projects through the gap. The sides of this projection are headwater curves, so the area grows wider the farther it projects. When the boundary hits the second river, it turns to follow it. It turns the acute angle, toward the other side of the projecting area. If both sides hit the same river, the boundary closes the end of the projection. But, if it reaches another headwater, it turns again and resumes following the grid. Both sides of the projecting area must be developed simultaneously to determine where the end closes. Projections of the zero-bridge area may extend quite far, but eventually all will close, and the colored line will form a continuous boundary across the map from side to side.

The area just beyond this first boundary can be reached on a direct route from the starting line by crossing only one river. The farther boundary of this one-bridge area can be constructed the same way as the first boundary, but using the first boundary as a starting line (figure 38). Using a second color, trace the river closest to the first colored boundary and bend around this river's headwater following the grid.

Though the second boundary is similar to the first, new situations arise. If the second river flows into the first river, the second boundary hits the first boundary. At this point,



figure 38

the second boundary must leave the river and follow the grid to form a confluence curve (see page 64). It continues along a line of the grid until it hits the third river, and then turns the acute angle. At no time can colored boundaries be superimposed.

If the first boundary arrives at the second river as a headwater curve, the second boundary will again hit the first. Here, too, the second boundary crosses the river and joins a grid line, continuing the same headwater curve by which the first boundary arrived. Though the second color may touch the first at a point, they must never coincide. When the ends of all its projections are closed, this second colored boundary will also be a single continuous line across the map from side to side.

Repeat this process as often as necessary to complete the map, using a new color for each new boundary. In principle only two colors are needed because all boundaries are continuous, but in practice the areas defined are often so narrow that six or eight colors

are needed to maintain clarity. They can be arranged in a rainbow that starts over after the last color. There is no need to assign numbers; the relative number of rivers can be ascertained by comparing colors, and the totals are irrelevant.

It is possible to fill in the entire map, but once shadow areas between routes become obvious they need no detail.

The first boundaries to reach the far edge define the local minimum number of river crossings in any direct route across the map. Every section of the finish line in which the number of crossings is lower than in either adjacent section is the end of a primary route area (see page 64). The areas that contain ridge routes appear as projections, with sawtooth sides like crude christmas trees, their bases on the far edge. The saw teeth contain dead ends that are good routes from the starting line, but do not remain good routes to the finish. The sawtooth edges can be trimmed off by working back across the map to the starting line.

Where colored boundaries reach the far edge, they are following grid lines. Here, they coincide with the boundaries of primary route areas. Tracing over them with a heavy dark pen constructs boundaries of route areas (figure 39). In the return direction, these boundaries are converging and the route area gets narrower. When the first river on the return pass is reached, the boundary from the first pass bends sharply and joins the river. The boundary of the route area does not bend as far. It turns to the other angle of the grid and follows a grid line, so the route area gets gradually wider. When the grid line hits another colored boundary, turn again, and follow the colored line. The boundaries of route areas constructed on the

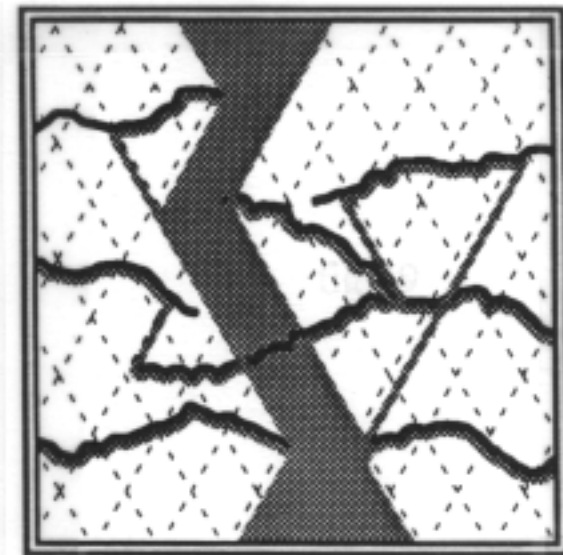


figure 39

second pass follow only lines of the grid and turn only 120 degree angles. They must never cross a boundary from the first pass, except when that boundary coincides with a river. If the route area boundary, following a colored line hits a river and the next colored line continues the same headwater curve, the route area boundary goes straight ahead. Using this method a contour map of travel cost can be constructed. Primary route areas show clearly on this map. Secondary routes can also be read, though sometimes only with difficulty.