SOLSTICE

Institute of Mathematical Geography

IMAGE

IMAGINATION

AN ELECTRONIC JOURNAL OF GEOGRAPHY AND MATHEMATICS

VOLUME II, NUMBERS 1 AND 2
1991

Institute of Mathematical Geography
2790 Briarcliff
Ann Arbor, MI 48105
IMaGe@UMICHUM; Solstice@UMICHUM

IMaGe

SOLSTICE

Institute of Mathematical Geography

IMAGE

IMAGINATION

Journal of the
Institute of Mathematical Geography

Summer, 1991

The purpose of *Solstice* is to promote interaction between geography and mathematics. Articles in which elements of one discipline are used to shed light on the other are particularly sought. Also welcome, are original contributions that are purely geographical or purely mathematical. These may be prefaced (by editor or author) with commentary suggesting directions that might lead toward the desired interaction. Individuals wishing to submit articles, either short or full- length, as well as contributions for regular features, should send them, in triplicate, directly to the Editor-in-Chief. Contributed articles will be refereed by geographers and/or mathematicians. Invited articles will be screened by suitable members of the editorial board. IMaGe is open to having authors suggest, and furnish material for, new regular features.

Send all correspondence to:

Institute of Mathematical Geography
2790 Briarcliff
Ann Arbor, MI 48105-1429
(313) 761-1231
IMaGe@UMICHUM

This document is produced using the typesetting program, TEX, of Donald Knuth and the American Mathematical Society. Notation in the electronic file is in accordance with that of Knuth's *The TEXbook*. The program is downloaded for hard copy for on The University of Michigan's Xerox 9700 laser- printing Xerox machine, using IMaGe's commercial account with that University.

Unless otherwise noted, all regular features are written by the Editor-in-Chief.

Upon final acceptance, authors will work with IMaGe to get manuscripts into a format well-suited to the requirements of *Solstice*. Typically, this would mean that authors would submit a clean ASCII file of the manuscript, as well as hard copy, figures, and so forth (in camera-ready form). Depending on the nature of the document and on the changing technology used to produce *Solstice*, there may be other requirements as well. Currently, the text is typeset using TEX; in that way, mathematical formulæ can be transmitted as ASCII files and downloaded faithfully and printed out. The reader inexperienced in the use of TEX should note that this is not a "what-you-see-is-what-you-get" display; however, we hope that such readers find TEX easier to learn after exposure to *Solstice*'s e-files written using TEX!

Copyright will be taken out in the name of the Institute of Mathematical Geography, and authors are required to transfer copyright to IMaGe as a condition of publication. There are no page charges; authors will be given permission to make reprints from the electronic file, or to have IMaGe make a single master reprint for a nominal fee dependent on manuscript length. Hard copy of *Solstice* is available at a cost of $15.95 per year (plus shipping and handling; hard copy is issued once yearly, in the Monograph series of the Institute of Mathematical Geography. Order directly from IMaGe. It is the desire of IMaGe to offer electronic copies to interested parties for free-as a kind of academic newsstand at which one might browse, prior to making purchasing decisions. Whether or not it will be feasible to continue distributing complimentary electronic files remains to be seen. Thank you for participating in this project focussing on environmentally-sensitive publishing.

ISBN: 1-877751-52-9

SUMMARY OF CONTENT

1. FULL-LENGTH ARTICLE.

Sandra L. Arlinghaus, David Barr, John D. Nystuen.

The Spatial Shadow: Light and Dark—Whole and Part

This account of some of the projects of sculptor David Barr attempts to place them in a formal, systematic, spatial context based on the postulates of the science of space of William Kingdon Clifford (reprinted in *Solstice*, Vol. I, No. 1).

2. REGULAR FEATURES

i. **Construction Zone** — The logistic curve

ii. **Educational feature** — Lectures on "Spatial Theory."

THE SPATIAL SHADOW: LIGHT AND DARK—WHOLE AND PART

"Life's but a walking shadow"
Shakespeare, *Macbeth*.

Sandra L. Arlinghaus, David Barr, John D. Nystuen

Introduction

Sunlight and shadow, day and night, solstice and equinox, lunar and solar eclipse–all are astronomical events that transform the surface of the earth into an event focused on the contrast between light and dark. The diurnal dynamics of the sweeping edge of the darkness are a foundation critical to the well-being of life on earth. Artistic expressions are numerous, ranging from Amish quilt patterns ("sunlight and shadow") to Indonesian shadow puppets. From a spatial standpoint, the mantle of night serves as a continuum linking disparate elements of the earth's surface; it is a whole composed of unseen parts.

WHOLE AND PART:
A Sculptural Unification of Unseen Parts.

The Four Corners Project

"The Four Corners Project," conceived in 1976 and completed in 1985, consists of an invisible tetrahedron spanning the inside of the earth with the four separate corners, made of marble, protruding from the crust of the earth (Figure 1). [1] These individual marble corner-markers (each about four inches high) were positioned in Easter Island, South Africa, Greenland, and New Guinea, with imaginary planes extending through the earth from each corner to the other three. The length of the imaginary line planned to link each pair of terrestrial markers is approximately 6465 miles. [2] One must know what a tetrahedron looks like and expand the scale of this knowledge to the scale of the entire earth to view this sculpture. In this respect, the art follows the pattern of the natural astronomical, global patterns of light and dark that require some sort of global perspective to envision a whole created from disparate unseen parts.

This tetrahedron is larger than proximate space. It is an abstraction that can be appreciated, as a whole, only in the mind; images of it created visually through written, printed, and verbal records encompass a broader view of it than does any collection of images taken from arbitrary physical vantage points in the universe. It is a shared perception, transcending language, that spans the minds of those who participate. [3] It requires abstract visualization, rather than physical vision, to "see" the entire sculpture.

This sculpture creates a conceptual unit from discrete parts that coalesces the evolutionary sequence of constructivistic, structurist art as well as the philosophical concerns of Zen gardens. In the structurist vocabulary, the art work draws the physical eye from one discrete component to another, and the unity of the work is revealed through the relationships of the components rather than through singular objects. In an early effort (1934), Henry Moore ("Four-Piece Composition") used the negative space of the sculpture to draw the physical eye, in proximate space, from one discrete component to another in order to suggest a single reclining figure [4]. The Zen garden at Ryoan-ji has stones arranged deliberately so that the whole can never be totally seen from a single perspective. Thus, the viewer, as in the Four Corners Project, must always be in a less than "divine" physical, perceptual position. Structurist reliefs emphasize the relationships among parts rather than the characterization of the parts themselves; [5] in this regard, "Four Corners" is a structurist concept at a global

Figure 1. The Four Corners Project. Four marble tetrahedra, each 4 inches high, mark the corners of a suggested, invisible, tetrahedron inscribed in the earth. Side length of the suggested large tetrahedron is about 6465 miles. Marker locations are in Easter Island, South Africa, Greenland, and New Guinea.

scale. In all of these cases, the unity of the entire piece unfolds naturally only when a leap of the imagination gives wholeness to the sculpture–whether that leap is in proximate or global space.

Geographical Background of the Four Corners

Barr fixed the general positions for the four corners on landmasses, using a globe and dividers; Nystuen pin-pointed each, using rotation matrices to align the North-South pole-based graticule with one using Easter Island and its antipodal point in the Thar Desert as poles. [6] Easter Island was chosen as the initial corner on account of its numerous cultural connections to the history of sculpture.

Embedding this tetrahedron in the earth-sphere (using the Clarke ellipsoid circumference of 24,873.535 miles [7]) required theoretical assumptions but also reflected the empirical facts of land/water distribution on earth--no corner was to be submerged in a lake or ocean. The environment and local surface materials surrounding the chosen corners are apt--from the igneous rock below a volcanic island, to the granitic sand in a desert, to the crystalline forms in an ice cap, to the organic material of a mangrove swamp. Indeed, the choice of the tetrahedron within the earth-sphere intentionally reflects the structure of the carbon atom as a fundamental component of life.

In 1980, Barr began to place the vertices of the tetrahedron; Table 1 shows the itinerary. The process that led to the completed product in 1985 involved the participation, from initial struggle to eventual respect and acceptance, of people from backgrounds not usually linked to the world of art: African veldt farmers, Eskimos, Irian Jayan missionaries, soldiers, police, politicians, and diplomats (for example, Table 1 shows the names of most of the airplane pilots who participated in the placement of these corners--they suggest the rich diversity of peoples associated with various aspects of this project).

TABLE 1
Log of travels associated with placement of the four corners
Listing compiled by Heather and Gillian Barr.

DESTINATIONS	**NAME OF AIRPLANE CAPTAIN**
DECEMBER AND JANUARY, 1980-81	
MACHU PICCHU, EASTER ISLAND, AND SOUTH AFRICA	
Detroit to Miami	John Bosh
Miami to Lima	Dick Rudman
Lima to Cuzco	Hugo Bisso
Cuzco to Lima	Eduardo Camino
Lima to Santiago	Javier Mesa
Santiago to Easter Island	Alphonso Estay
Easter Island to Santiago	Gustavo Vila
Santiago to Buenos Aires	Sergio Kurth
Buenos Aires to Cape Town	Carlos Bustamante
Cape Town to Johannesburg	Steev Kaup
Johannesburg to New York	Tony Laas
New York to Detroit	Hal Grenddin

JULY, 1981

GREENLAND

Windsor to Montreal	Mr. Golze
Montreal to Frobisher	Mr. Savage
Frobisher to Sonderstrom Fjord	Patty Doyle
Sonderstrom Fjord to Ice Cap	Patty Doyle
Ice Cap to Sonderstrom	Patty Doyle
Sonderstrom to Godthab (Nuuk)	Patty Doyle
Godthab to Frobisher	Sven Syversen
Frobisher to Montreal	Carl Gitto
Montreal to Windsor	Louis Ghyrmothy

JANUARY, 1985

IRIAN JAYA

Djajpura to Danau Bira	Poambang Kuncaro a.k.a."Bang Bang Koon"
Danau Bira to Djajpura	Bang Bang Koon
Djajpura to Biac	Mr. Fujiono
Biac to Ujung Pandang	Mr. Darynato
Ujung Pandang to Bali	Angus Tiansyah
Bali to Djakarta	Mr. Sunarto
Djakarta to Singapore	Mr. Tan

In December of 1980, Barr and his party (which included other fine artists and a professional dancer) went to Machu Picchu, where the tetrahedral marble pinnacles were washed at the ancient ceremonial site (at the sundial called "ini-huatana" ("hitching post of the sun")), prior to placement in the ground. From there they went to Easter Island; surveying equipment of William Mulloy [8], a member of Thor Heyerdahl's expedition to that island, was used to place the first vertex of the tetrahedron on January 4, 1981 (Table 2), one minute of longitude from the calculational center of 109° 25'30". This location has elevation just above sea level and is in a former leper colony.

TABLE 2
Geographic coordinates of the Four Corners

Site	Latitude	Longitude
Easter Island	27°06′20″ S	109°24′30″ W
South Africa	27°30′36″ S	024°06′00″ E
Greenland	72°38′24″ N	041°55′12″ W
New Guinea [actual]	02°20′50″ S	138°00′00″ E
New Guinea [planned]	02°06′36″ S	137°23′24″ E

imperfections are imperceptible. It is only with our imaginations that we can appreciate the difference between the ideal and terrestrial forms.

Mathematical Uniqueness of the Four Corners —Extensions of the idea

When spherical trigonometry was applied to a map showing all landmasses whose antipodal points are also land-based (Figure 2) [10] it was possible to prove that the choice of a tetrahedron as a shape for this sculpture is unique within the set of regular polyhedra called "Platonic" solids. [11] Plato linked the set of five regular polyhedra (tetrahedron, cube, octahedron–polyhedron with eight triangular faces, dodecahedron–polyhedron with twelve pentagonal faces, and icosahedron–polyhedron with twenty triangular faces) with five basic components from which he believed the earth to have been formed. [12] No Platonic solid, other than the tetrahedron, can be embedded in the earth with all corners on land, one of which is on Easter Island. [13]

It also follows from the mathematics that, although the tetrahedron is unique as a choice, there are an infinite number of possible positions in which it might have been oriented within the earth (Figure 3). The possibilities for the corners other than Easter Island are, however, tightly constrained within the arcs of the circle of "latitude" (centered on C, the antipodal point of Easter Island in the Thar Desert) shown in Figure 3. (An azimuthal equidistant projection was used because distances measured from the center are true.) Once a point is chosen within one of these arcs as a corner site, the choices for the other two corners are forced (as the remaining vertices of an equilateral triangle inscribed in the circle of "latitude"). [14] These three sites form the triangular base of a tetrahedron with Easter Island (unseen in Figure 3) at the apex of the solid, on the other side of the earth from the center of the circle in Figure 3.

The after-the-fact discoveries that the choice of the tetrahedron was unique within the set of Platonic solids, and that the extent of infinite "play" in site selection could be constrained within specified bounded intervals, enhance the planned selection of Easter Island as the choice for the initial vertex of the tetrahedron. Indeed, other choices were considered as an initial vertex; however, the idea of using this tiny patch of land in the Pacific hemisphere as the anchor for this "titanic" tetrahedron of terrestrial sites, not only proved possible, but irresistible as well.

LIGHT AND DARK: **A problem of boundary.**

Natural boundaries, such as those between water and land, are often crenulated and complex. Many words are necessary to translate a natural boundary into a cadastral sur-

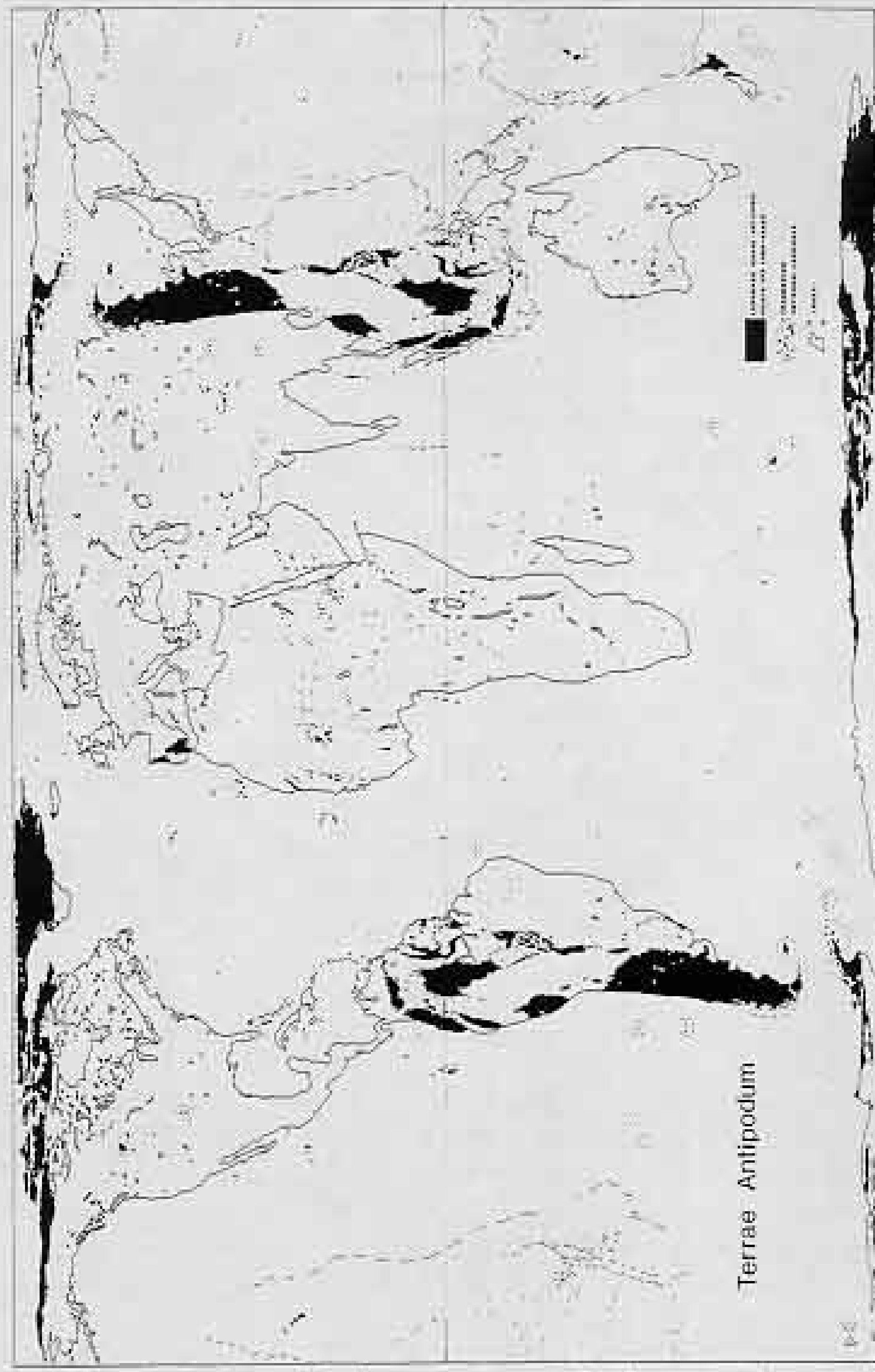


Figure 2. Terrae Antipodum. Dark areas represent landmasses whose antipodal points are on land. Fragmented antipodal landmasses (archipelagos) are encircled by dashed lines. Antipodal continental outlines are shown (where needed to understand the map) over the ocean as dashed lines. The base map is a Peters projection. The equator bisects the vertical neat line. This map was used to establish uniqueness of the choice of a tetrahedron within Barr's constraints.

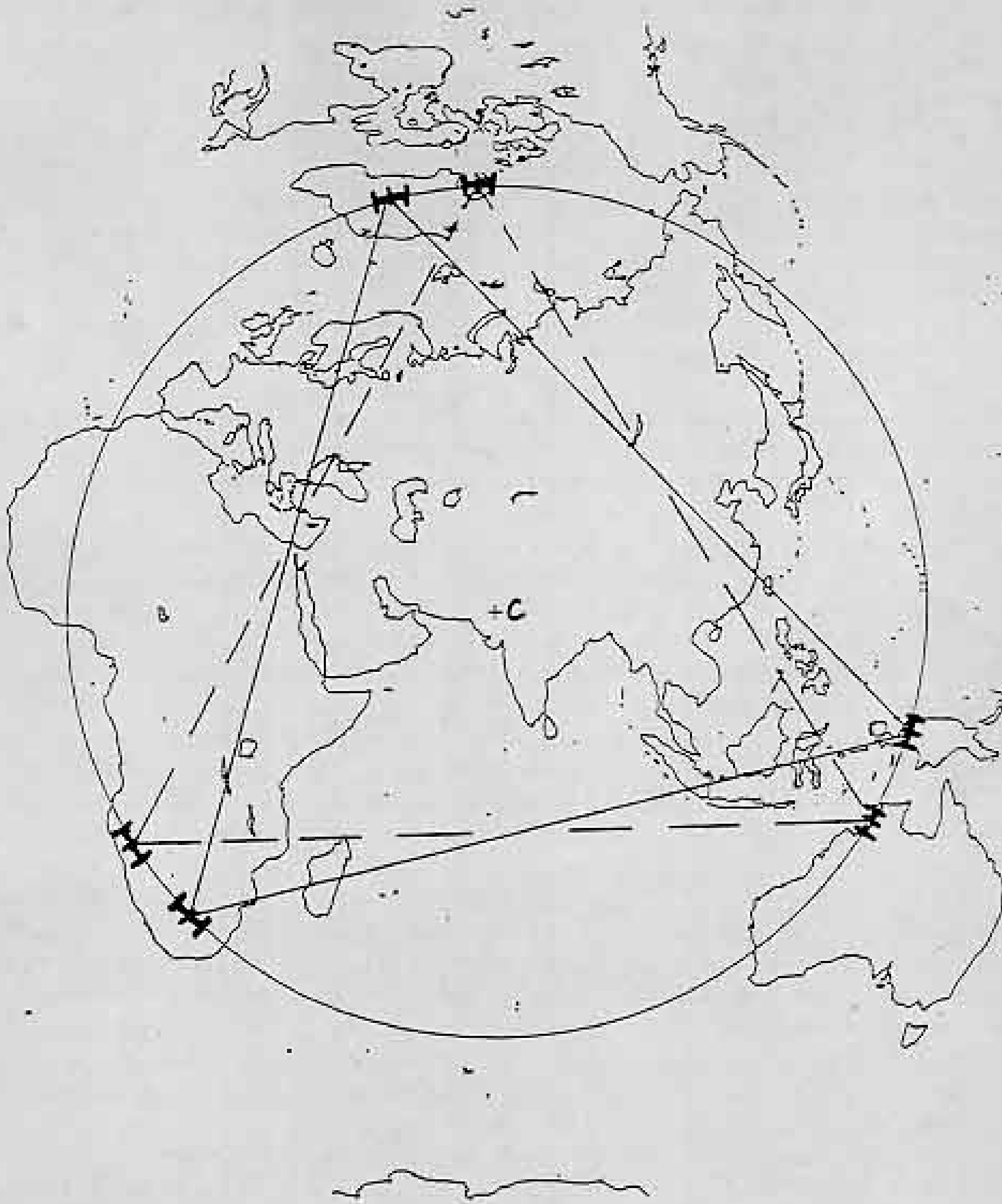


Figure 3. Shaded intervals show all possible land-based locations for three corners of the base of a tetrahedon with Easter Island as apex of the solid inscribed in the earth. Easter Island is antipodal to the center of the circle, C. The base map is an azimuthal equidistant projection. Any distance measured from the center, C, is true.

vey description. At places where the abstract and natural boundaries intersect, interesting arrangements can arise.

SunSweep

SunSweep is a sculpture in three separate locations along the U.S./Canadian border that was designed to commemorate the peaceful interaction across this border. Its three parts are located at places where natural and abstract boundaries intersect. The western terminus is on a bit of U.S. territory which can only be reached, on land, by passing through Canada. The eastern terminus is on a bit of Canadian territory which can only be reached, on land, by passing through the United States. Thus, a nice symmetry is created by the intersection of a natural and an abstract boundary; this symmetry is intentionally reflected in the choices for the locations and in the physical shapes of the elements of the SunSweep sculpture (Figure 4). The sculpture represents the arch of the sun in the sky from east to west. Coincidentally, perhaps, Barr noted a common social outlook among the people inhabiting these anomalous locations– they appeared to share a kind of independence coming from this blurred boundary, suggesting a unity in social perspective associated with this sculpture.

Geographical Background of *SunSweep*

The eastern-most piece, arching inland, is situated on Campobello Island in New Brunswick; the western-most piece, also arching inland, is on Point Roberts in the State of Washington; and, the keystone of the arch, composed of two separate stone elements, is on an island in the Lake-of-the-Woods in Minnesota (Figure 5: a, b, c). Each piece is about five feet tall and is formed from selectively polished flame-finished black Canadian granite carved, in Michigan, from one mass.

These markers that trace the sweep of the sun across the celestial sphere were sited close to the U.S./Canadian border to commemorate the spirit of cooperation between these two countries. A hand print, suggesting "I was here," has been lasered into the polished stone–a "Canadian" print on one side pressing against its mirror image "United States" print on the other side.

The choice of locations for the sculpture suggests the path of the sun; they were selected with an eye to displaying the interplay of ideas between astronomical sweep and political boundary–as geographic "boundary dwellers" in the world of art. [15] They were also selected for their characteristic of physically forcing (in terms of access) interdigitation between U.S. and Canadian boundaries.

Thus Campobello Island, maintained as an International Park, is the site for the eastern piece; the arch is situated on Ragged Point (Table 3), a Canadian location accessible by road only through the United States. The trail leading to the sculpture is the "SunSweep" Trail, formerly known as the "Muskie Trail" and re-named at the suggestion of Senator Edmund Muskie of the State of Maine. The western-most piece of the arch is situated in Lighthouse Park on Point Roberts (Table 3), a United States community at the southern tip of a spit of land that is accessible (by land) only through Canada. American Point (Penasse Island), Minnesota, the northernmost U.S. island (Table 3) in the Lake-of-the-Woods (Lake situated on the U.S./Canadian border), is close to a U. S. peninsula which is accessible by land only through Canada; it is the site of the keystone for the arch in the locale referred to as "Northwest Angle" which, other than those in Alaska, contains the only U.S. landmasses north of the 49th parallel.

Figure 4. SunSweep. The 5-foot high earth-markers set out on a lawn, prior to placement along the U.S./Canada border.

Grooves lasered into the sides of one element of the keystone piece and the top edge of the sculpture offer visitors the opportunity to tie location to selected astronomical events. The top edge is angled so that a sunbeam is parallel to it on the summer solstice; a groove in one side is angled to align with the sun on both equinoxes; and, a groove on the other

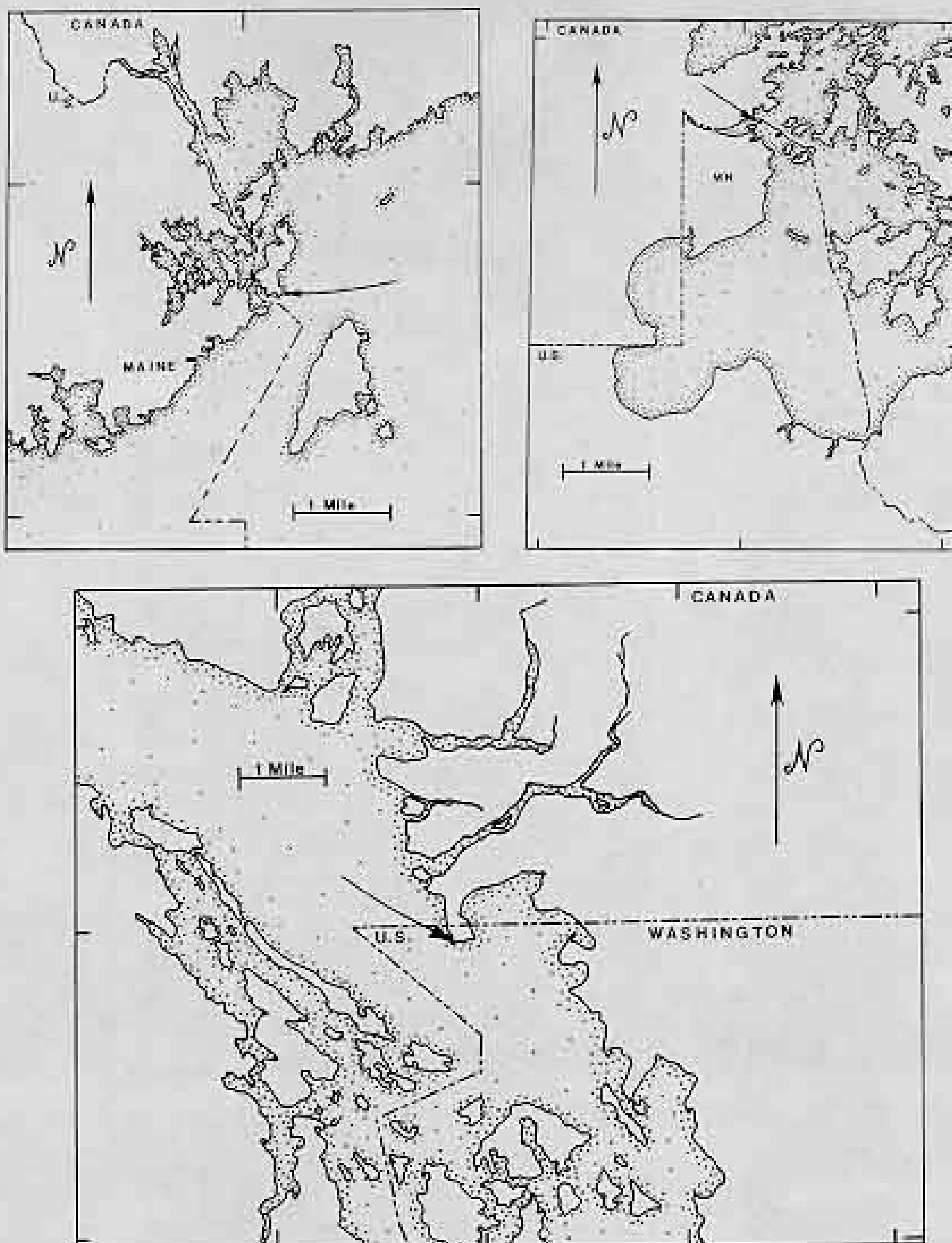


Figure 5. a, b, c. Maps of the three SunSweep sites (a, New Brunswick; b, Minnesota; c, Washington) emphasizing interdigitation associated with anomalous locations along the U.S./Canada boundary.

side is angled to align with the sun on the winter solstice. The shadows cast by a sunbeam at each astronomical event would suggest a tracing on the ground, with the succession of the seasons, in the shape of an analemma [16], calling to mind the equation of time and

TABLE 3
Geographic coordinates of SunSweep

Site	Latitude	Longitude
Campobello Island, NB	44°50′10″ N	066°55′25″ W
Point Roberts, WA	48°58′23″ N	123°05′00″ W
Lake-of-the-Woods, MN	49°21′45″ N	094°57′40″ W

ultimately Kepler's Laws of planetary motion. [17]

The second element of the Minnesota piece is aligned to the North Star. These markers were installed on the summer solstice of 1985. The alignments to the sun on this date and to the North Star appeared true. The pieces in New Brunswick and Washington were aligned subtly to each solstice and equinox position using the beveling planes of the granite and the orientation in the pattern of sited, smaller rocks surrounding the sculpture.

The markers at each site have a bronze plaque set in the concrete base describing their metaphor. At the installation of the sculpture in Washington, the arch arrived broken and was cemented together as it was set into concrete in the ground. [18] Future generations who come across this irregular crack might wonder what it "means," and whether or not it represents an alignment to some peculiar astronomical event. At best, it might be regarded as a remnant of a transportation system not geared to shipping heavy, brittle items with great success! The local citizenry is reconciled to the crack and in fact take delight in this sculpture as their "Liberty Bell."

Mathematical Extensions of the ideas behind *SunSweep*

These three locations, selected initially for unique boundary characteristics, closely approximate ideal geometric placement along an arc of a great circle. A summary of how the actual measurements differ from the "ideal" ones is shown in Table 4. The keystone location is, in fact, not halfway between the ends as one might hope for in a perfect arch. The great circle distance from the New Brunswick site to the Minnesota site is longer than the distance from the Minnesota site to the Washington site.

TABLE 4
Great Circle Distances between SunSweep Sites.

Sites	Distance in miles
Campobello Island to Lake-of-the-Woods	1347
Lake-of-the-Woods to Point Roberts	1263
[SUM:	2610]
Campobello Island to Point Roberts	2605
[Mid-point of entire great circle sweep	1302.5]

In addition, the three locations, as a set, do not lie along a single great circle; ideally, it might have been desirable to have them do so in order to keep the arch within a single plane passing through the earth's center. This sort of ideal arrangement was not possible, however, because of the requirement of interdigitation of U.S. and Canadian boundaries. Still, the actual placement of the markers is quite close to the ideal: the great circle distance from the New Brunswick location to the Washington location is 2605 miles–only 5 miles shorter than the sum of the component distances. Indeed, the midpoint of the great circle arc joining the New Brunswick location to the Washington location is at about 49 degrees 5 minutes North Latitude, 93 degrees 56 minutes West Longitude–a great circle distance of about 60 miles to a site east and slightly south of the actual location of the sculpted keystone. As was the case with "Four Corners," the unity of the entire "SunSweep" unfolds naturally only when a leap of the imagination gives wholeness to the sculpture; in this case that wholeness is suggested by a sequence of anomalous locations along a political boundary.

Political boundaries are abstract and often simply defined, an advantage in conflict resolution. The "Oregon Question" that agitated England and the United States for a generation was resolved during the James Polk administration (1846) by the simple agreement to extend the northwestern boundary along the 49th parallel from the Lake-of-the-Woods to the Pacific, [19] an arc of 1263 miles (great circle distance 1256 miles). Vancouver Island extends south of this line but the continental boundary ends where the 49th parallel reaches Puget Sound.

The fact that the great circle distance between the western and middle sculpture sites rounds off to the same length as the length of the U.S./Canadian land border along the 49th parallel was unplanned in the sculpture. As was the case with the uniqueness of the choice of the tetrahedron for the Four Corners Project, this too was an after-the-fact discovery, linking both geography and mathematics to sculpture.

THE SPATIAL SHADOW:

A theoretical framework.

The emergence of the after-the-fact discoveries surrounding these sculptures suggests the suitability of looking for theory to link the concepts underlying these particular art projects, much as poetry might be after-the-fact theory linking already- existing word-images. To do so, we draw on the interdisciplinary ties linking mathematics to geography, and linking both to art.

Thus, we adopt a view in which mathematics includes the science of abstract space; in which geography ties this science of space to the real world; and, in which art offers abstract means to appreciate these ties. A set of postulates of the "science of space" were created in the late nineteenth century by William Kingdon Clifford drawing only on commonsense notions of continuity and discreteness, flatness, magnification and contraction, and similarity, that formed part of the foundation of the non-Euclidean geometries at the base of modern physics. [20] By considering a set of fundamental relations, simply expressed, it became possible to analyze spatial relations in a fashion that did not rely solely on Euclid's postulates, and particularly not necessarily on Euclid's parallel postulate. [21]

We consider a transformational approach to theory, echoing the emphasis of contemporary "global" mathematics in seeking properties which remain invariant when carried via transformation from one space to another. It might be tempting to consider sunlight as a basic unit, because light coming through the sculpture is what links the geometry of the sculpture with the reality of the earth. With the sun at an "infinite" distance from earth, its beams are parallel to each other (from our vantage point). Incoming solar radiation might therefore be considered an "affine" transformation (in which sets of parallel lines are invariant) that maps elements protruding from the earth's surface as shadows onto the earth's surface (as in a structurist relief). [22] There are a number of appealing elements to this particular transformational approach. The affine transformation is the basis of much computer software for displaying graphics, suggesting a natural alignment of theory and computer mapping in order to merge the mathematics of sculptural structures with the spatial relations of the earth. [23]

Because such an approach has the concept of affine transformation at its heart, however, it necessarily emphasizes the notion of parallelism. Our emphasis is, rather, on separate pieces whose relationship creates a single unit of art composed of separate parts intentionally devoid of interest in order to focus on that relationship, as (quotation attributed to Einstein)

"History [Art] consists of relationships rather than events"
A. Einstein.

It seems therefore, inappropriate to forge a linkage with theory based on parallelism. Far more suitable is to follow the lesson learned from Clifford and find basic elements that better match that which we seek to characterize. [24]

The concept of shadow, rather than the affine transformation that creates the shadow, seems a better choice as a fundamental unit with which to work. Single spatial shadows (of physical objects) are discrete units of individual character; yet, they change in response to diurnal fluctuations, eventually to become united in a single nighttime continuum under the global spatial shadow of the earth on itself. Indeed, the concept of shadow, itself, also embodies the notion of transformation–

"The shadows now so long do grow,
That brambles like tall cedars show,
Molehills seem mountains, and the ant
Appears a monstrous elephant."
Charles Cotton, *Evening Quatrains*.

"Shadow" is dynamic mathematically, as a transformation, as well as geographically, as the sweeping boundary separating light from dark that refreshes the earth on a daily cycle. Shadow is tied directly to time through the diurnal motions of the Earth, and it is tied indirectly to time, at a personal level, as well. Each individual casts a personal time-shadow–a long trail of experiences representing accumulated wisdom over a period of years (and growing longer all the time), together with a short extension into a "cone" of opportunity, generated by a space-time continuum, into the near future. [25] The analysis of the manner in which these temporal shadows might become unified in some global manner [26] is no doubt better left to philosopy and religion as

"Time watches from the shadow".
W. H. Auden, Birthday Poem.

With spatial shadows and temporal shadows, one might recast Clifford's postulates for a Science of Space as Postulates for light and dark based on the concept of shadow. The contrast between light and dark, and sunlight and shadow, gives insight into the shape of things; or, as Clifford put it,

"Out of pictures, we imagine a world of solid things,"

a statement reminiscent of Plato's "Den". [27] That is, a shadow is a creature that exists as a transformation of a three dimensional object onto a two-dimensional surface much as the relief format is the transitional step from two-dimensional paintings to full three dimensional art. The shape and position of the shadow are a function of

1. the shape of the three-dimensional object,
2. the orientation of the three-dimensional object in relation to the light source, and
3. the curvature of the receiving surface.

The concept of shadow links these elements and therefore represents a relationship that is "structurist" in nature.

Clifford's statement of his postulates for a Science of Space follows. [28]

"1. Postulate of Continuity. Space is a continuous aggregate of points, not a discrete aggregate.
2. Property of Elementary Flatness. Any curved surface which is such that the more you magnify it, the flatter it gets is said to possess elementary flatness.
3. Postulate of Superposition. A body can be moved about in space without altering its size or shape.
4. Postulate of Similarity. According to this postulate, any figure may be magnified or diminished in any degree without altering its shape."

Both "space" and "darkness" are diffuse, rather than linear, as concepts; their "lateral" character suggests that they, and other concepts possessing this characteristic, such as time, continuity, or inclusion/exclusion, have the power to unify. Thus, we rethink Clifford's postulates within his stated context, to see if they can be reasonably recast as a different set of postulates concerning light and dark.

Shadow Postulates

1. Postulate of Continuity. Total darkness is a continuous aggregate of shadow, and not a discrete aggregate of individual shadows.

Indeed, total darkness on the earth is continuous as it is formed from a single global shadow of the earth on itself; all other shadows are lesser. This global shadow is a limiting position that a sum of discrete aggregates of shadow might approach but never reach; the whole is greater than the sum of its parts.

2. Postulate of Equinox. On every surface which has this property, all but a finite number of points are such that they are in darkness and light an equal amount of time.

Clifford notes that any surface that possesses his property of elementary flatness is one on which "the amount of turning necessary to take a direction all round into its first position is the same for all points on the surface." This is suggestive of what happens on earth at the time of the equinoxes in which all parallels of latitude are bisected by the edge of darkness so that all but the poles spend half the diurnal cycle in light and half in dark. Hence the restatement of "Elementary Flatness" as "Equinox."

3. Postulate of Unique Position. The length and angle of individual shadows impart information, in a unique fashion, as to position on earth.

One consequence of Clifford's Postulate of Superposition is that "all parts of space are exactly alike." A body can be moved about in space without altering its size or shape, but its shadow changes at every different location on earth (at a given instant). Thus the Postulate of Unique Position is parallel to that of Superposition.

4. Postulate of Solstice. On every surface which has this property, all but a finite number of points are such that they are in darkness and light an unequal amount of time.

Using the idea in Clifford's Postulate of Similarity, any shadow of a single object may become magnified or diminished in any degree, through time. However, the shape of the object which casts the shadow remains unchanged. The Earth's shadow always covers exactly half of the earth-sphere (in theory). The dark/light boundary slips over the Earth's surface covering half of it in darkness, altering the extent to which shadows of unchanged objects become magnified or diminished. During this process, not all points experience the same amount of darkness. Hence, "Similarity" is replaced with "Solstice." The dynamics of this process are bounded between two parallels (the Tropics), so that there is also implied parallelism associated with this Postulate, just as Lobatschewsky noted implied parallelism associated with Clifford's fourth postulate and rejected it in order to consider using his geometry to understand astronomical space. [29]

Now this set of postulates "fits" with the earth and its shadow (indeed, the earth motivated it). The reader wishing to determine where the dark/light boundary appears at a given time at a given location need only perform the following construction, [30] using a globe on a sunny day. Point the north pole of the globe toward the earth's north pole (make compensating adjustments for southern hemisphere locations), where meridians of longitude converge. Rotate the globe on this north/south axis until your location appears on top of the globe–where a plane "parallel" to the surface of the earth is tangent to the globe. The shadow cast by the sun on the globe will trace out accurately the position of the light/dark boundary on the earth at that moment. This construction works because it amounts to putting the globe in exactly the position that the earth is in relative to the sun–it is a good example of Shadow Postulate 3 concerning Unique Position because the globe position required is unique for each point on earth, even though each unique position will generate the same position for the shadow. (Postulate 1 applies, and Postulate 2 applies on two days of the year and Postulate 4 applies otherwise.)

A natural next step is then to turn these postulates back around on the style of sculpture (that of discrete units that suggest unity) that motivated them. Shadow is a sort of underlying, continuous and rhythmical, [31] phrasing in a poetry of dark and light. The postulates offer a strategy to see what "poetic images" can be formed within this poetic phrasing.

SunSweep is a sculpture in three discrete parts. Thus, Shadow Postulates 1, 2, and 4, which are tied to continuity are not of particular interest, though they are significant in explaining the sun-sighting from each position. Shadow Postulate 3, dealing with Unique Position, is the natural, abstract "line" of logic joining the sites, as the "Sunsweep." Light coming through the keystone is what merges its geometry with the reality of the earth, as a seasonal analemma traced out on the earth by pencils of sunlight. The concept of light and dark, viewed within the concept of Unique Position, is what abstractly links the three SunSweep sites, and their sun-sighting capability, as a unit.

With the Four Corners Project, we have the possibility of considering the more global postulates because of the requirement of a global view from which to visualize the entire sculpture. In this case, the interesting alignments of sculpture with theory appear to be in the Equinox and Solstice Shadow Postulates. Four Corners may be referenced using standard geographic latitude and longitude, but it is most easily referenced using a spherical coordinate system of latitude and longitude based on a polar axis through one of the four corners and its antipodal point. Rotation matrices, from linear algebra, may then be used to move from one coordinate system to the other. Thus, if one views the Four Corners Project as having a

"North" Pole at the Greenland corner, it seems natural to ask whether or not "Equinox" and "Solstice" relative to this coordinate system coincide with astronomical equinox and solstice positions of the earth. Indeed, the concepts apply, but the results are different.

Because the only parts of this earth-scale sculpture touched by sunlight are the corners: "equinox" occurs when exactly two of the corners are illuminated and two are in the earth's shadow; "solstice" occurs otherwise. "Equinox" is clearly a more frequent occurrence with the Four Corners than it is with the Earth. In this view, the natural concept drawing the Four Corners together as a unit is that of spatial relations between Earth and the Solar System as Equinox and Solstice, and at the same time, this human construct of "Four Corners" enlightens the natural occurence of equinox and solstice.

In both cases, the postulates of light and dark serve as a natural abstract line to suggest unity, much as the physical positioning of proximate discrete pieces suggests natural lines along which to sight in a wide range of artistic efforts. This is an alignment of fundamental ideas. It is reasonable to consider therefore where this might lead, both in terms of art and in terms of formal theory.

Further directions appear two-fold: first, in the world of art, it may be useful to consider other existing art in this after-the-fact mode and then to employ these postulates as part of a plan in developing discrete sculpture to suggest unity; and second, in the world of formal theory, it seems appropriate to extend abstract theory from the postulates with an eye to possibly turning it back around on art. One direction that is currently being investigated by Kenneth Snelson is in the arena of mathematics applied to spheres, particularly to those applications developed in analogy with the earth's position in the solar system. Pauli's Exclusion Principle of quantum mechanics, which rests on likening the spin of an electron to the diurnal spinning of the earth on its axis, serves as a sort of spatial starting point for his alignments of modern physics and sculpture. [32] (According to Pauli's principle, no two electrons can be in the same orbit of the nucleus. [33]) In a related, but different, direction, the use of Clifford's postulates suggests that a suitable extension of ideas might arise in the world of various non-Euclidean geometries and particularly in those whose Euclidean models are often cast in terms of a sphere.

Notes

1. David Barr, 1979. "Notes on celebration," *The Structurist*, 8: pp. 52-56. David Barr, 1981, "The four corners of the world." *Coevolution* 5: 5. David Barr, 1982, "The four corners project," *Museum Catalogue*, Meadow Brook Art Gallery.
2. For published documentation of Nystuen's original calculations, estimated originally by Barr, see Sandra L. Arlinghaus and John D. Nystuen, *Mathematical Geography and Global Art: the Mathematics of David Barr's 'Four Corners Project'*. (Ann Arbor: Institute of Mathematical Geography, 1986), Monograph #1.
3. Susan Ager, 1984. "It's a titanic tetrahedron." *Detroit Free Press*, Sunday, July 15: A-1, A-11. Marsha Miro, 1985. "David Barr's amazing cosmic art adventure." *Detroit Free Press*, Sunday Magazine Section. October 6: 6-12; 18-22. Smithsonian Institution Documentary Film. 1986. "In celebration: David Barr's Four Corners Project." Archives of American Art. Released in April at The Detroit Institute of Arts.
4. James J. Sweeney, *Henry Moore*, (New York: The Museum of Modern Art, 1946), p. 31.
5. Eli Bornstein, "The search for continuity in art and connectedness with nature," *The Structurist: Continuity and Connectedness*, No. 29/30, 1989-90, pp. 38-45.
6. Ibid., all of note 2.
7. Simo H. Laurila, *Electronic Surveying and Navigation*. (New York: Wiley, 1976).
8. William Mulloy (Ph.D.) late Professor of Archaeology, University of Wyoming.
9. Szykula, D. and Dwaihy, E. The Four Corners Dance. 1976- 1985.
10. Waldo R. Tobler, 1961, "World map on a Moebius strip." *Surveying and Mapping*, 21:486. Sandra L. Arlinghaus, "Terrae antipodum" in *Essays on Mathematical Geography — II*, (Ann Arbor: Institute of Mathematical Geography, 1987), Monograph #5.
11. Ibid., reference, only, note 2.
12. Hermann Weyl, *Symmetry*, (Princeton: Princeton University Press, 1952). Athelstan Spilhaus, 1975. "Geo-art: tectonics and Platonic solids." *Transactions of the American Geophysical Union*, 56. No. 2. Plato, "Timaeus."
13. Ibid., note 9.
14. Ibid.
15. John D. Nystuen, 1967, "Boundary shapes and boundary problems." *Papers of Peace Research Society International* 7:107- 128.
16. Erwin Raisz, 1941, The analemma. *The Journal of Geography*. 40:90-97.
17. Johannes Kepler, *Prodromus Dissertationem Mathematicarum continens Mysterium Cosmographicum*, (Tubingen, 1596).
18. Grace Productions. 1986. "Time Is No Object." Videotape, premiering on Michigan Public Television, January 15, 1989.
19. J. D. Hicks, *The Federal Union: A History of the United States to 1865*. (Cambridge, MA: The Riverside Press, 1948).
20. William Kingdon Clifford, 1990 (reprint). The postulates of the science of space. *Solstice: An Electronic Journal of Geography and Mathematics*, Vol. I, No. 1, pp. 6-16. Coxeter,

H. S. M. 1965. *Non-Euclidean Geometry*. (Toronto: University of Toronto Press, 1965) pp. 8-11.

21. Ibid., Clifford.

22. Sandra L. Arlinghaus, "Solar Woks," in *Essays on Mathematical Geography-II*, (Ann Arbor: Institute of Mathematical Geography, 1987), Monograph #5,

23. Sandra L. Arlinghaus, 1990, "Fractal geometry of infinite pixel sequences: super-definition resolution?" *Solstice: An Electronic Journal of Geography and Mathematics*, Vol. 1, No. 2, 48-53.

24. Saunders Mac Lane, "Proof, Truth, and Confusion" The 1982 Ryerson Lecture, The University of Chicago, (Chicago: The University of Chicago Press, 1982).

25. Albert Einstein, (reprint). *Relativity: The Special and the General Theory*. (New York: Bonanza Books, 1961).

26. Henry Moore and W. H. Auden, *Auden Poems, Moore Lithographs: An Exhibition of a Book Dedicated by Henry Moore to W. H. Auden with Related Drawings*. (London: The British Museum, 1974).

27. Plato's *Republic*. The Complete and Unabridged Jowett Translation. (New York: Airmont Publishing, 1968), pp. 267-272, end book VI, beginning book VII.

28. Ibid., note 21.

29. Ibid., note 20.

30. William W. Bunge. Personal communication to John D. Nystuen.

31. Richard D. Cureton, *Rhythmic Phrasing in English Verse*. (London: Longmans, 1991 (in press)). Andrea Voorhees Arlinghaus, 1991, personal communication to Sandra L. Arlinghaus.

32. Encyclopaedia Brittanica. "Quantum Mechanics," (Chicago: William Benton, 1966). Vol. 18, p. 929. Kenneth Snelson, *The Nature of Structure*. (New York: New York Academy of Sciences, 1989). pp. 21-24.

33. Ibid., Snelson.

Sandra L. Arlinghaus is Director, Institute of Mathematical Geography, 2790 Briarcliff, Ann Arbor, MI 48105

David Barr is Professor of Art, Macomb Community College, South Campus, Warren, MI 48093

John D. Nystuen is Professor of Geography and Urban Planning, The University of Michigan, Ann Arbor, MI 48109

Written July, 1990, revised, March, 1991, and June, 1991. All funds for the art projects described herein were supplied by David Barr.

Construction Zone

Simple analysis of the logistic function

A derivation supplied by S. Arlinghaus in response to questions from William D. Drake, School of Natural Resources, University of Michigan, concerning aspects of his interest in transition theory. Discussed Tuesday, May 6, 1991, Colloquium in Mathematical Geography, IMaGe. Present: Sandy Arlinghaus, Bill Drake, John Nystuen (this commentary is included in *Solstice* at the request of the latter).

1. The exponential function–unbounded population growth

Assumption: The rate of population growth or decay at any time t is proportional to the size of the population at t.

Let Y_t represent the size of a population at time t. The rate of growth of Y_t is proportional to Y_t;

$$dY_t/dt = kY_t$$

where k is a constant of proportionality.

To solve this differential equation for Y_t, separate the variables.

$$dY_t/Y_t = k\,dt; \int 1/Y_t\,dY_t = \int k\,dt.$$

Therefore,

$$\ln|Y_t| = kt + c_0.$$

Consider only the positive part, so that

$$Y_t = e^{kt+c_0} = e^{c_0}e^{kt}.$$

Let $Y_{t0} = e^{c_0}$. Therefore,

$$Y_t = Y_{t0}\,e^{kt};$$

exponential growth is unbounded as $t \longrightarrow \infty$.

Suppose $t = 0$. Therefore,

$$Y_t = Y_{t0}\,e^0 = Y_{t0}.$$

Thus, Y_{t0} is the size of the population at $t = 0$, under conditions of growth where $k > 0$ (Figure 1).

2. The logistic function–bounded population growth.

Assumption appended to assumption for exponential growth. In reality, when the population gets large, envirnomental factors dampen growth.

The growth rate decreases– dY_t/dt decreases. So, assume the population size is limited to some maximum, q, where $0 < Y_t < q$. As $Y_t \longrightarrow q$ it follows that $dY_t/dt \longrightarrow 0$ so that population size tends to be stable as $t \longrightarrow \infty$. The model is exponential in shape initially and includes effects of environmental resistance in larger populations. One such algebraic expression of this idea is

$$dy_t/dt = kY_t \cdot \frac{q - Y_t}{q}$$

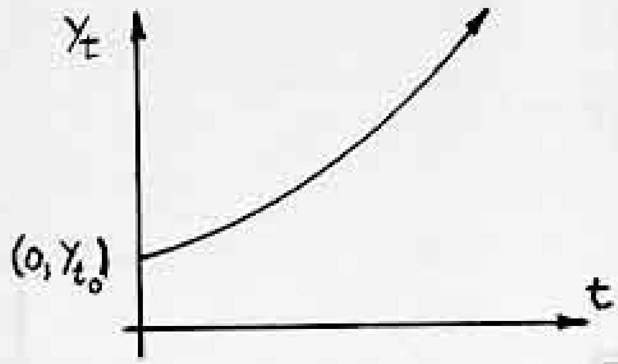


Figure 1

because in the factor $\frac{q-Y_t}{q}$, when Y_t is small, $\frac{q-Y_t}{q}$ is close to 1 (and the growth therefore close to the exponential) and when Y_t is close to q, $\frac{q-Y_t}{q}$ is close to 0, and the growth rate dY_t/dt tapers off. This factor acts as a damper to exponential growth.

Replace k/q by K so that

$$dY_t/dt = KY_t(q - Y_t)$$

and the rate of growth is proportional to the product of the population size and the difference between the maximum size and the population size.

Solve this latter equation for Y_t: (separate the variables)

$$\frac{dY_t}{Y_t(q - Y_t)} = K\,dt; \int \frac{dY_t}{Y_t(q - Y_t)} = \int K\,dt$$

Use a Table of Integrals on the rational form in the left-hand integral:

$$\frac{1}{q}\ln\left|\frac{Y_t}{q - Y_t}\right| = Kt + C$$

$$\ln\left|\frac{Y_t}{q - Y_t}\right| = qKt + qC.$$

Because $Y_t > 0$ and $q - Y_t > 0$,

$$\ln\frac{Y_t}{q - Y_t} = qKt + qC.$$

Therefore,

$$\frac{Y_t}{q - Y_t} = e^{qKt+qC} = e^{qKt}e^{qC}.$$

Replace e^{qC} by A. Therefore,

$$\frac{Y_t}{q - Y_t} = Ae^{qKt};$$
$$Y_t = (q - Y_t)Ae^{qKt};$$
$$Y_t = qAe^{qKt} - Y_tAe^{qKt};$$
$$Y_t(Ae^{qKt} + 1) = qAe^{qKt};$$
$$Y_t = \frac{qAe^{qKt}}{Ae^{qKt} + 1};$$

now divide top and bottom by Ae^{qKt}, equivalent to multiplying the fraction by 1, so that

$$Y_t = \frac{q}{1 + \frac{1}{Ae^{qKt}}} = \frac{q}{1 + \frac{1}{A}e^{-qKt}}.$$

Replace $1/A$ by a and $-qK$ by b producing a common form for the logistic function (Figure 2),

$$Y_t = \frac{q}{1 + ae^{bt}}$$

with $b < 0$ because $b = -qK$, and $q,\ K > 0$.

3. Facts about the graph of the logistic equation.

a. The line $Y_t = q$ is a horizontal asymptote for the graph.

This is so because, for $b < 0$,

$$\lim_{t \to \infty} \frac{q}{1 + ae^{bt}} \longrightarrow \frac{q}{1 + a(0)} = q$$

Can the curve cross this asymptote? Or, can it be that

$$Y_t = \frac{Y_t}{1 + ae^{bt}}?$$

Or,

$$1 = 1 + ae^{bt}?$$

Or,

$$ae^{bt} = 0$$

Or, that $a = 0$? No, because $a = 1/A$. Or, that $e^{bt} = 0$–no.

Thus, the logistic growth curve described above cannot cross the horizontal asymptote so that it approaches it entirely from one side, in this case, from below.

b. Find the coordinates of the inflection point of the logistic curve.

Vertical component:

The equation $dY_t/dt = KY_t(q - Y_t) = KqY_t - KY_t^2$ is a measure of population growth. Find the maximum rate of growth–derivative of previous equation:

$$d^2Y_t/dt^2 = Kq - 2KY_t$$

To find a maximum (min), set this last equation equal to zero.

$$Kq - 2KY_t = 0$$

Therefore, $Y_t = q/2$. This is the vertical coordinate of the inflection point of the curve for Y_t, the logistic curve–dY_t/dt is increasing to the left of $q/2$ ($d^2Y_t/dt > 0$) and dY_t/dt is decreasing to the right of $q/2$ ($d^2Y_t/dt < 0$). So, the maximum rate of growth occurs at $Y_t = q/2$. [The rate at which the rate of growth is changing is a constant since the first differential equation is a quadratic (parabola)].

Horizontal component:

To find t, put $Y_t = q/2$ in the logistic equation and solve:

$$q/2 = \frac{q}{1 + ae^{bt}}$$

Solving,

$$1 + ae^{bt} = 2; e^{bt} = 1/a; e^{-bt} = a; -bt = \ln a,$$

$$t = \frac{\ln a}{-b}$$

Thus, the coordinates of the inflection point of the logistic curve are:

$$(\ln a/(-b), q/2).$$

In order to track changes in transitions, such as demographic transitions, monitoring the position of the inflection point might be of use. To consider feedback in such systems, graphical analysis (Figure 2) of curves representing transitions might be of use.

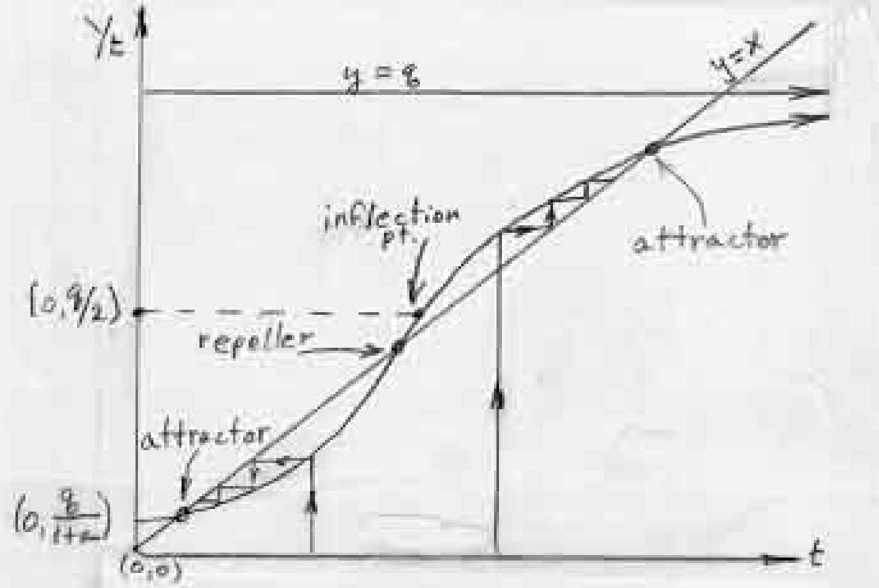


Figure 2. The intersection points of the line $y = x$ with the logistic curve are, using terms from chaos theory, attractors on either end, and a repelling fixed point in the middle, possibly near the inflection point of the curve.

Educational Feature
Topics in Spatial Theory
Based on lectures given by S. Arlinghaus
as a guest speaker in John Nystuen's
Urban Planning, 507, University of Michigan
Feb. 21, 28, 1990; four hours

The people along the sand
All turn and look one way.
They turn their back on the land.
They look at the sea all day.

...

They cannot look out far.
They cannot look in deep.
But when was that ever a bar
To any watch they keep?

Robert Frost *Neither Out Far Nor In Deep*

I. Introduction

Theory guides the direction technology takes; mathematics is the theoretical foundation of technology. To become more than a mere user of various software packages and programming languages, which change rapidly (what is trendy in today's job market may be obsolete tomorrow), it is therefore critical to understand what sorts of decisions can be made at the theoretical level. Underlying theory is "spatial" in character, rather than "temporal," when the objects and processes it deals with are ordered in space rather than in time (most can be done in both–decide which is of greater interest). The focus with GIS is spatial; hence, the theory underlying it is "spatial."

This is not a new idea; D'Arcy Thompson, a biologist, saw (as early as 1917) a need for finding a systematic, theoretical organization of biological species that went beyond the classification of Linnaeus. What he found to be fundamental, to characterization along structural (spatial, morphological) lines (rather than along temporal, evolutionary lines) was the "Theory of Transformations"–in Thompson's words:

"In a very large part of morphology, our essential task lies in the comparison of related forms rather than in the precise definition of each; and the deformation of a complicated figure may be a phenomenon easy of comprehension, though the figure itself have to be left unanalysed and undefined. This process of comparison, of recognising in one form a definite permutation or deformation of another, apart altogether from a precise and adequate understanding of the original 'type' or standard of comparison, lies within the immediate province of mathematics, and finds its solution in the elementary use of certain method of the mathematician. This method is the Method of Co-ordinates, on which is based the Theory of Transformations. [*The mathematical Theory of Transformations is part of the Theory of Groups, of great importance in modern mathematics. A distinction is drawn between Substitution-groups and Transformation-groups, the former being discontinuous, the latter continuous–in such a way that within one and the same group each transformation is infinitely little different from another. The distinction among biologists between a mutation and a variation is curiously analogous.]*

I imagine that when Descartes conceived the method of co-ordinates, as a generalisation from the proportional diagrams of the artist and the architect, and long before the immense possibilities of this analysis could be foreseen, he had in mind a very simple purpose; it was perhaps no more than to find a way of translating the form of a curve (as well as the position of a point) into numbers and into words. This is precisely what we do, by the method of coordinates, every time we study a statistical curve; and conversely translate numbers into form whenever we 'plot a curve', to illustrate a table or mortality, a rate of growth, or the daily variation of temperature or barometric pressure. In precisely the same way it is possible to inscribe in a net of rectangular co-ordinates the outline, for instance, of a fish, and so to translate it into a table of numbers, from which again we may at pleasure reconstruct the curve.

But it is the next step in the employment of co-ordinates which is of special interest and use to the morphologist; and this step consists in the alteration, or deformation, of our system of co-ordinates, and in the study of the corresponding transformation of the curve or figure inscribed in the co-ordinate network.

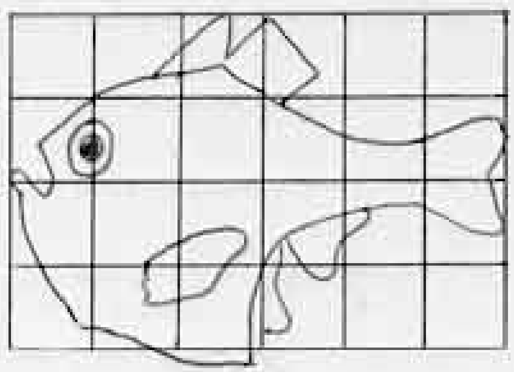
Let us inscribe in a system of Cartesian co-ordinates the outline of an organism, however complicated, or a part thereof: such as a fish, a crab, or a mammalian skull. We may now treat this complicated figure, in general terms, as a function of x, y. *If we submit our rectangular system to deformation on simple and recognised lines, altering, for instance, the direction of the axes, the ratio of* x/y, *or substituting for* x *and* y *some more complicated expressions, the we obtain a new system of co-ordinates, whose deformation from the original type the inscribed figure will precisely follow. In other words, we obtain a new figure which represents to old figure under a more or less homogeneous strain, and is a function of the new co-ordinates in precisely the same way as the old figure was of the original co-ordinates* x *and* y.

The problem is closely akin to that of the cartographer who transfers identical data to one projection or another [reference below]; and whose object is to secure (if it be possible) a complete correspondence, in each small unit of area, between the one representation and the other. The morphologist will not seek to draw his orgainc forms in a new and artificial projection; but, in the converse aspect of the problem, he will enquire whether two different but more or less obviously related forms can be so analysed and interpreted that each may be shown to be a transformed representation of the other. This once demonstrated, it will be a comparatively easy task (in all probability) to postulate the direction and magnitude of the force capable of effecting the required transformation. Again, if such a simple alteration of the system of forces can be proved adequate to meet the case, we may find ourselves able to dispense with many widely current and more complicated hypotheses of biological causation. For it is a maxim in physics that an effect ought not to be ascribed to the joint operation of many causes if few are adequate to the production of it.

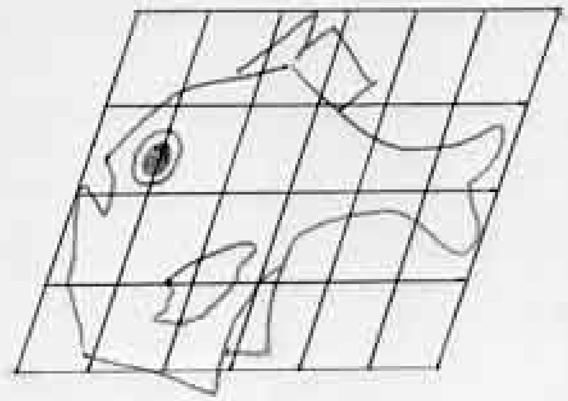
Reference: Tissot, Mémoire sur la representation des surfaces, et les projections des cartes géographiques (Paris, 1881)."

Sir D'Arcy Wentworth Thompson, pp. 271-272, in *On Growth and Form.*

Look at Thompson's comments concerning biological structure to see what parallels there are, already, with GIS structure and to see what they might suggest–compare to Tobler's map transformations.



I.1



I.2

Figure I Sample of Thompson's Transformations. Fig. I.1: Argyropelecus olfersi. Fig. I.2: Sternoptyx diaphana.

1. GIS (the digitizer) uses coordinates to translate forms (maps) into numbers.

2. All GIS software translates numbers into maps, which may then be printed out, parallel to inscribing a fish in a set of coordinates, translating it into a set of numbers, from which the fish may be reproduced at any time (Figure I.1)

3. Thompson's deformations correspond to the ideas of scale shifts on maps. Transformations describe shifts in scale. Figure I.2.

4. Thompson's comments on the distinction between discontinuous and continuous reflects partitioning of mathematics into discrete and continuous. Discrete need not be finite–look at two different types of garbage bag ties–twist ties and slip-through ties, and imagine them to be of infinite extent.

5. We see simple transformations in GIS–maps might be stretched or compressed in the vertical direction. Imagine using a small digitizing table to encode a large map by deliberately recording "wrong" positions—then use a transformation within the computer to correct the "wrong" positions so that the map prints out correctly on the plotter. Large digitizing tables become unnecessary.

6. We look, for future direction, to the Theory of Groups. For today, we confine ourselves to a few simple transformations.

II. Transformations

Transformations can allow you to relate one form to another in a systematic manner allowing retieval of all forms. To do this, you need to know how to define a transformation so that this is possible. Beyond this, one might consider a stripped-down transformation, for even more efficient compression of electronic effort [Mac Lane].

A. Well-defined (single-valued).

Let "tau" be a transformation carrying a set X to a set Y: in notation, $\tau : X \longrightarrow Y$.

Tau is said to be well-defined if each element of X corresponds to exactly one element of Y. Visually, this might be thought of in terms of lists of street addresses: the set X consists of house addresses used as "return" addresses on letters. The set Y consists of other house addresses. The transformation is the postal transmission of a letter from locations in X to locations in Y. A single value of X maps to single value of Y.

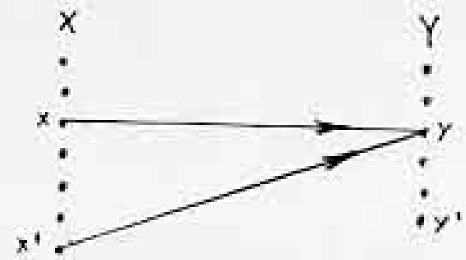


Figure II.1 This is a transformation–two distinct letters (x and x') can be posted to the same address (y). (Many-one map).

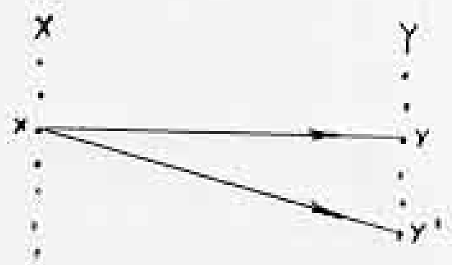


Figure II.2 This is NOT a transformation–one letter (x) cannot, itself, go to two different addresses (y and y') (new technology of e-mail permits this–suggests for possible need for change in fundamental definitions). (One-many map).

B. Reversible

i. One-to-one correspondence.

A one-to-one correspondence is a transformation in which each x in X goes to a distinct y in Y; the situation depicted in Figure II.1 cannot hold. From the standpoint of reversibility, this is important; if the situation in II.1 could hold how would you decide, in reversing, whether to "return" y to x or to x'??

ii. Transformations of X onto Y

A transformation of X onto Y is such that every element in Y comes from some element of X; there are no addresses outside the postal system (Figure II.3).

Figure II.3 This is a transformation–it is neither one-to-one, nor onto (y' is outside the system).

iii. A transformation τ from X to Y is reversible– it has an inverse τ^{-1} from Y to X if τ is one-to-one and onto; it has an inverse from a subset of Y to X if τ is one-to-one (Figure II.4).

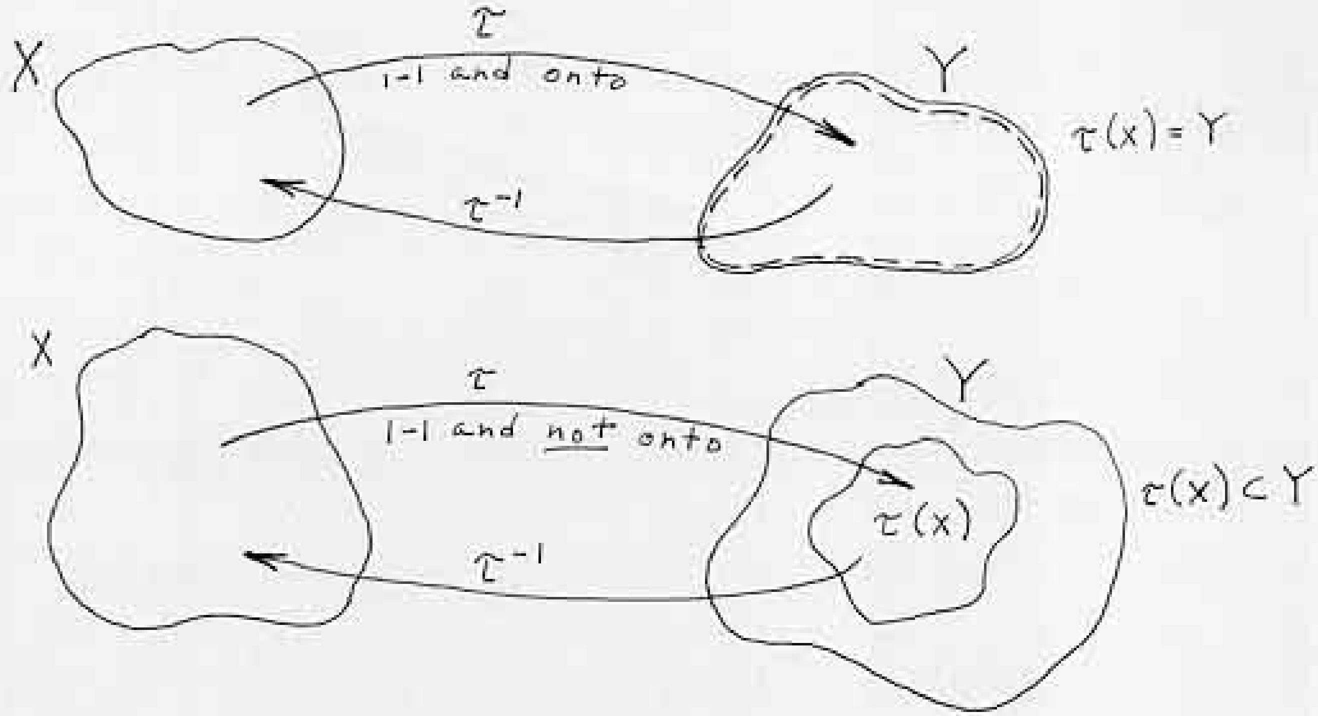


Figure II.4 In the top part, $\tau(X) = Y$. In the bottom part $\tau(X)$ is properly contained in Y; this is like data compression–like ZIP followed by UNZIP.

C. Rubbersheeting

The use of transformations that have inverses is critical in rubbersheeting; associations between data sets must be made in a manner so that correct information can be gained from the process.

III. Types of Transformations

One might consider moving objects within a fixed coordinate system, or holding the objects fixed and moving the coordinate system. Thompson did the latter; rubbersheeting does the latter; NCGIA materials (Lecture 28) comment that the latter approach is particularly well-suited to GIS purposes.

Two major types of transformations:

a. Affine transformations: these are transformations under which parallel lines are preserved as parallel lines. That is, both the concept of "straight line" and "parallel" remain; angles may change, however.

There are four types of affine transformations as noted on suitable NCGIA handout (Figure III.1). Products of affine transformations are themselves affine transformations.

Current technology employs types 1 and 2, quite clearly. CRT allows for translation of maps, and for scale change in y-direction only. Copier also allows for the same, and in addition, permits different shifts in scale along the two axes, allowing maps with different scales along different axes to be brought to the same scale and pieced together. (See output from Canon Color Copier.) On that output, the x-axis if fixed by the transformation and the y-axis is stretched to 200% of the original. Thus, a circle transforms to an ellipse, a rectangle with base parallel to the x-axis transforms to a larger rectangle, and a rectangle with base not parallel to the x-axis transforms to a parallelogram with no right angles (Figure III.2).

B. Curvilinear transformations; neither straightness nor parallelism is necessarily preserved (Thompson fish, Figure III.3).

IV. Exercise, page 5, lecture 28, NCGIA.

V. Steiner networks

If centers of gravity are used as a centering scheme in a triangulated irregular network, then it is desired to have no centroid lie outside a triangular cell. Thus, no cell should have angle greater than 120 degrees, so that the Steiner network (where all angles are exactly 120 degrees) will serve as an outer edge (a limiting position) for the set of acceptable triangulations. Thus, it is important to know how to locate Steiner networks.

VI. Digital Topology

The notion of a "triangulation" is a fundamental concept in topology (sometimes called "rubber sheet" geometry). "Digital" topology is a specialization of "combinatorial" topology in which the fundamental units are pixels. The same "important" theorems underlie each. The Jordan Curve Theorem (which characterizes the difference between the "inside" and the "outside" of a curve, is an example of such a theorem). Using concepts from digital topology, "picture" processing (as a parallel to "data" processing) is possible. There are numerous references in this field; some include works by geographer Waldo Tobler and by mathematician Azriel Rosenfeld. Other key-words to topics of interest in this area include,

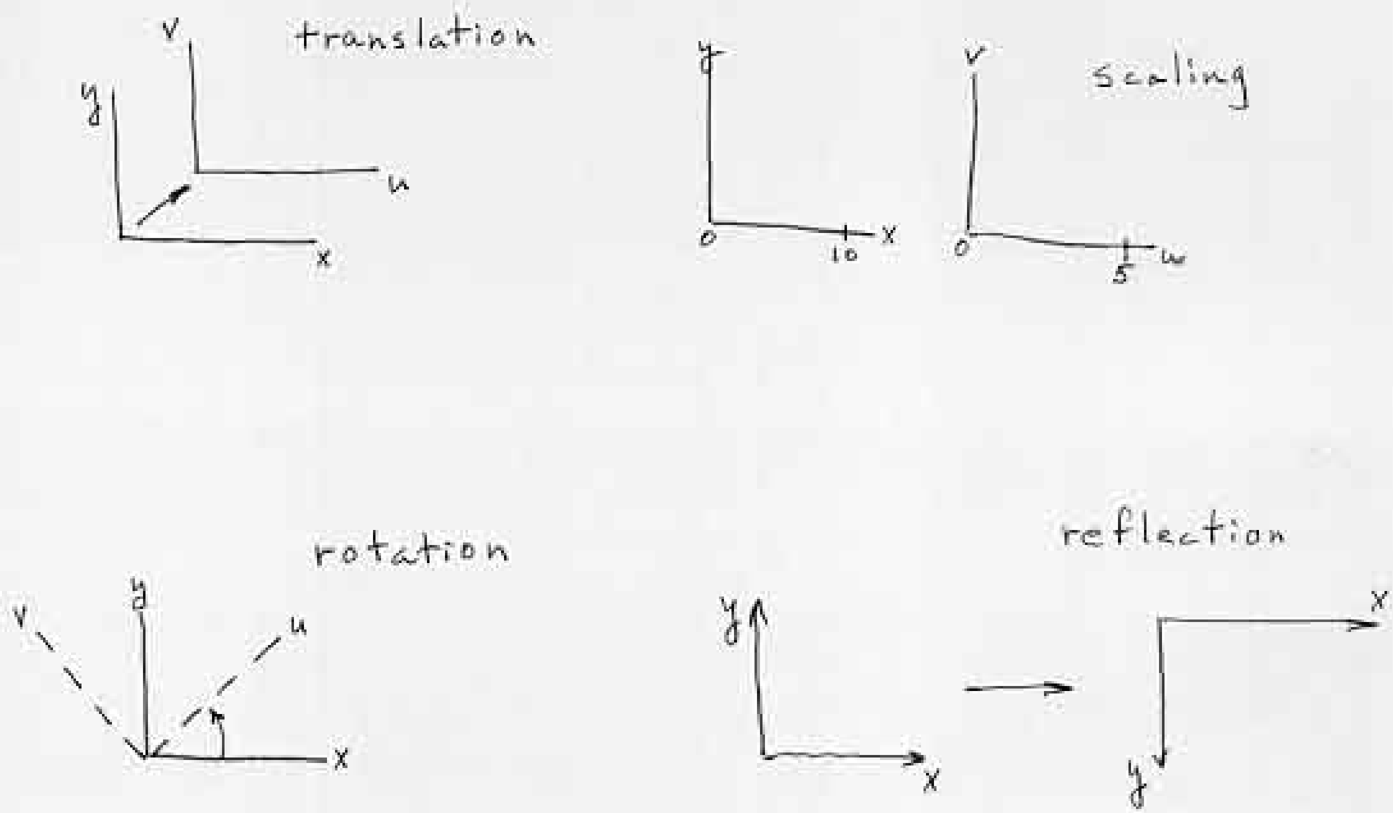


Figure III.1

Jordan Curve Theorem in higher dimensions; quadtrees; scale-free transformations; close-packings of pixels.

VII. The algebra of symmetry–some group theory

D'Arcy Thompson commented that the theory of transformations was tied to the theory of groups. A "group" is a mathematical system whose structure is simpler than that of the number system we customarily use in the "real-world." In our usual number system, we have two distinct operations of "+" and "x"; thus, we have rules on how to use each of these operations, and rules telling us how to link these two operations (distributive law; conventions regarding order of operations).

A group is composed of a finite set of elements, $S = \{a, b, c, \ldots, n\}$ that are related to each other using a single operation of "$*$." Under this operation, the set obeys the following rules (and is, by that fact, a group).

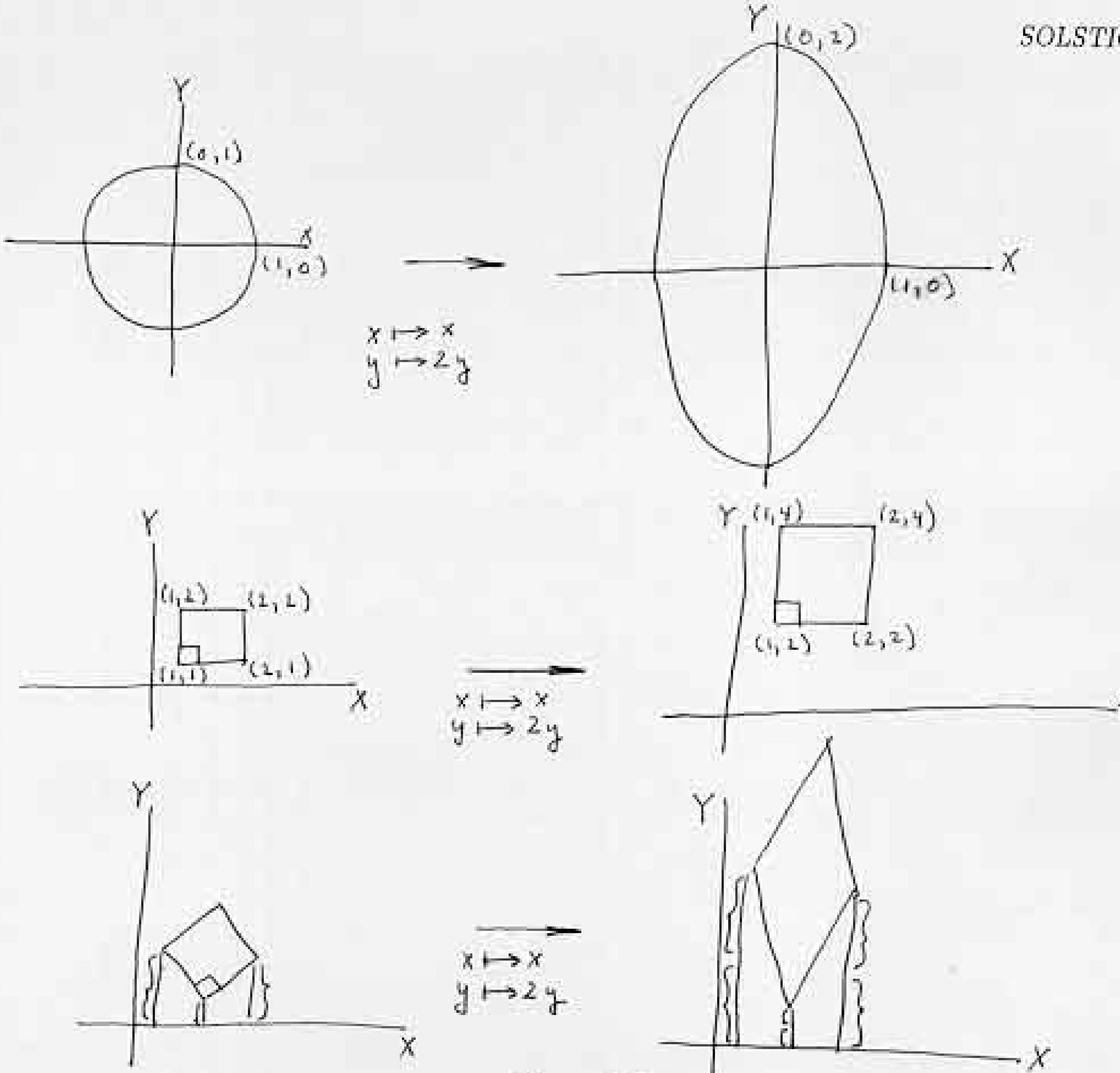


Figure III.2

1. The product, under $\star$, of any two elements of S is once again an element of S—this system is "closed" under the operation of $\star$—no new element (information) is generated.
2. Given a, b, and c in S: $(a \star b) \star c = a \star (b \star c)$. The manner in which parentheses are introduced is not of significance in determining the answer (information content) resulting from a string of operations under $\star$. The operation of $\star$ is said to be associative.
3. There is an identity element, 1, in S such that for any element of S, say a, it follows that

$$a \star 1 = 1 \star a = a.$$

4. Each element of S has an inverse in S; that is, for a typical element a of S, there exists another element, b of S, such that

$$a \star b = b \star a = 1.$$

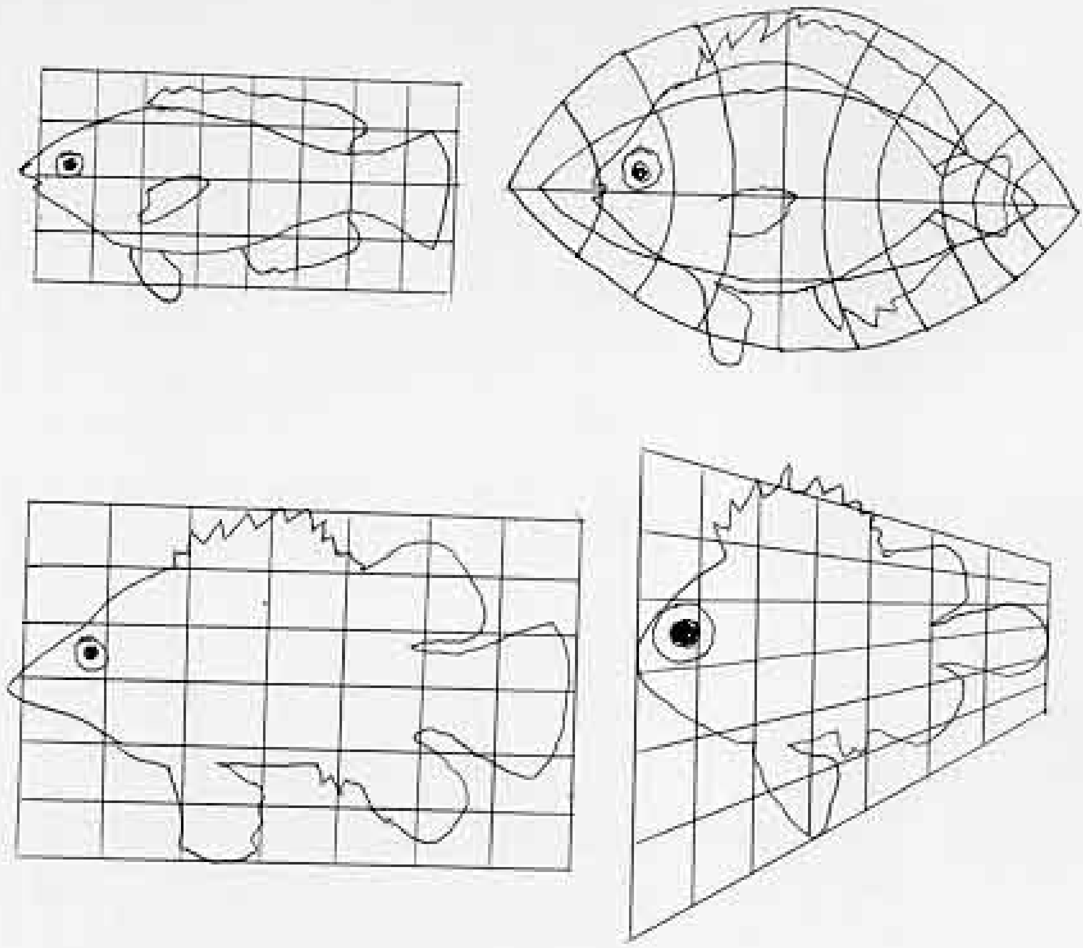


Figure III.3

Denote the inverse of a as a^{-1}. Thus, $a \star a^{-1} = a^{-1} \star a = 1$.

The order in which elements are related to each other, using $\star$, may matter; it need not be true that $a \star b = b \star a$. (Elements of the group do not necessarily "commute" with each other.)

The algebraic idea of "closure" is comparable to the GIS notion of snapping a polygon shut, so that chaining of line segments does not continue forever—the system is "closed."

A. The affine group; affine geometry.

The definition of group given above was to a set of elements and an operation linking them. These elements might be regarded as transformations. In particular, consider the set of all affine transformations of the plane that are one-to-one (translations, scalings, rotations, and reflections). These form a group, when the operation $\star$ is considered as the composition of functions:

i. The product of two affine transformations is itself an affine transformation;

ii. In a sequence of three affine transformations, it does not matter which two are grouped first, as long as the pattern of the three is unchanged—associativity.

iii. The affine transformation which maps the plane to itself serves as an identity element.

iv. Because the affine transformations dealt with here are one-to-one, they have inverses (all translations have inverses; only those linear transformations with inverses are considered here).

Affine geometry is the study of properties of figures that remain invariant under the group of one-to-one affine transformatons. Here are some theorems from affine geometry.

i. Any one-to-one affine transformation maps lines to lines.

ii. Any affine transformation maps parallel sets of lines to parallel sets of lines.

iii. Any two triangles are equivalent with respect to the affine group.

To demonstrate the theorem in iii., consider a fixed triangle with position (OB_0C_0), relative to an x/y coordinate system. Choose an arbitrary triangle, (ABC). Use elements of the affine group to move (ABC) to coincide with (OB_0C_0): a translation slides A to O (Figure VII.1). Two separate scaling operations and rotations slide B to B_0 and C to C_0. This is possible because O, B, and C are not collinear (as vectors, OB and OC are linearly independent).

This is the theoretical origin of the GIS notion that control points must be non-collinear and that there must be at least three of them. From a mathematical standpoint, it does not, therefore, matter whether the control points are chosen close together or far apart; however, from a visual standpoint it does matter. When control points are chosen close together the scaling operation required to transform the control triangle into other triangles is generally enlargement. When the control triangle is chosen with widely spaced vertices, the scaling operations required to transform it into other triangles is generally reduction. Errors are more visible with enlargement. Therefore, it is better, for the sake of visual comfort, to rely on reduction (reducing error size, as well) whenever possible, and therefore, to choose widely-spaced control points.

This is like the exercise above; there are two scalings and another affine transformation (here a translation, in the exercise, a reflection). In either case, the outcome of applying a sequence of affine transformations is still an affine transformation. In this case, it does not matter in what order the scaling operations are executed and in what order, relative to the scaling, the translation is applied. In the case of the exercise, however, this is not the case.

It does not matter in what order the scalings are applied. It is the case that $\tau_1 \circ \tau_2 = \tau_2 \circ \tau_1$. It is also the case that $\tau_1 \circ \tau_3 = \tau_3 \circ \tau_1$. However, it is not the case that

$$\tau_2 \circ \tau_3 = \tau_3 \circ \tau_2 :$$

$$(50,5) \xrightarrow{\tau_2} (50,48) \xrightarrow{\tau_3} (50,432)$$

$$(50,5) \xrightarrow{\tau_3} (50,475) \xrightarrow{\tau_2} (50,4660)$$

Observe, however, that it is possible to solve the problem applying the reflection earlier. Take τ_1 to be the required reflection so that y is sent to $50 - y$ (reflection before the scale

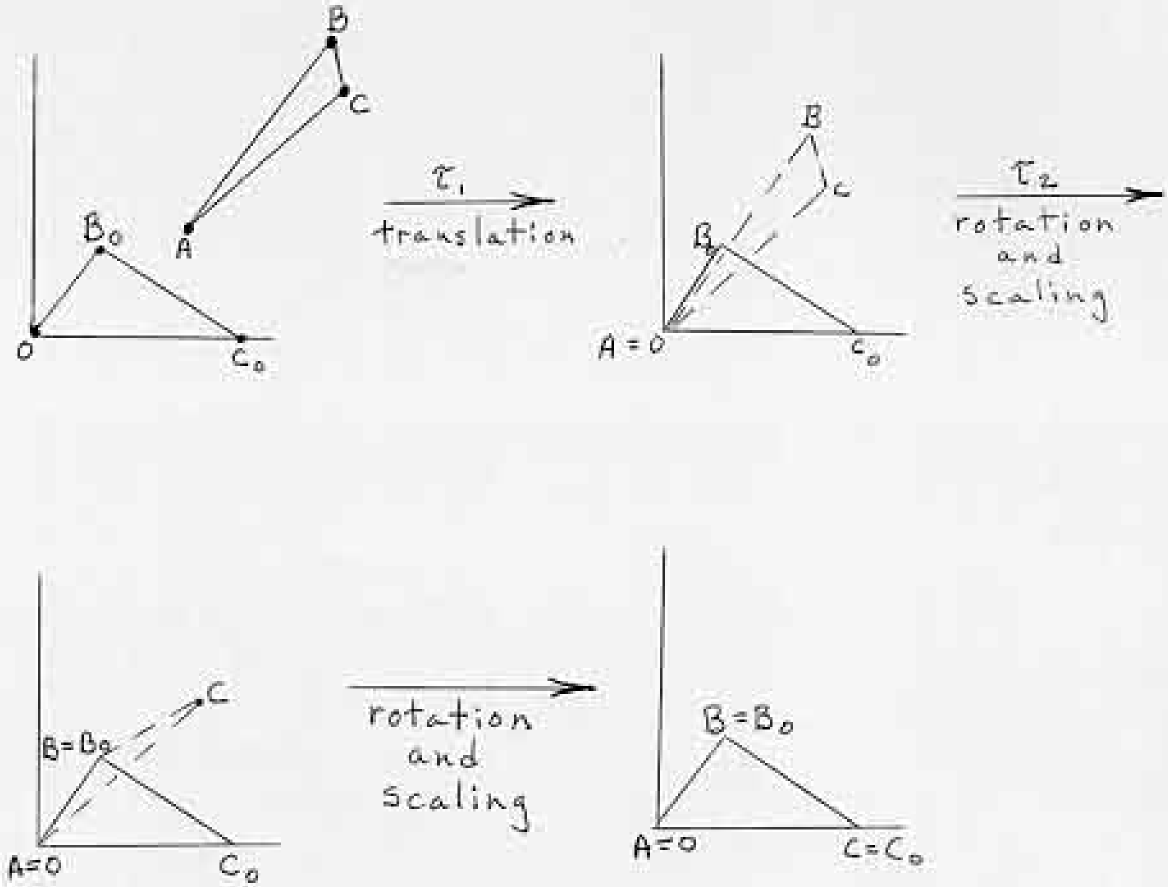


Figure VII.1

change on the y-axis). Figure VII.2 shows the solution here. In the non-commutative case here, there is a sharp difference in the "correct" y-value and the other possible one. In this case, as in the previous one, it does not matter how the application of transformations are separated by parentheses, and it is guaranteed that the product will itself be affine.

Thus, the order of application of affine transformations, within the group (locally), is important. This might cause difficulties (sending you off the screen), or it might be turned to an advantage in zooming-in on something. What caused the problem here was the reflection. Products of rotations of the plane are roatations of the plane; products of translations are translations, and products of scalings are scalings. Here, and as we shall see later, reflections cause non-commutativity (similar problems might have arisen in Figure VII.1, had a reflection been involved).

iv. Any triangle is affine-equivalent to an equilateral triangle (choose whatever control triangle desired—can choose an underlying lattice of regularly spaced triangular points and

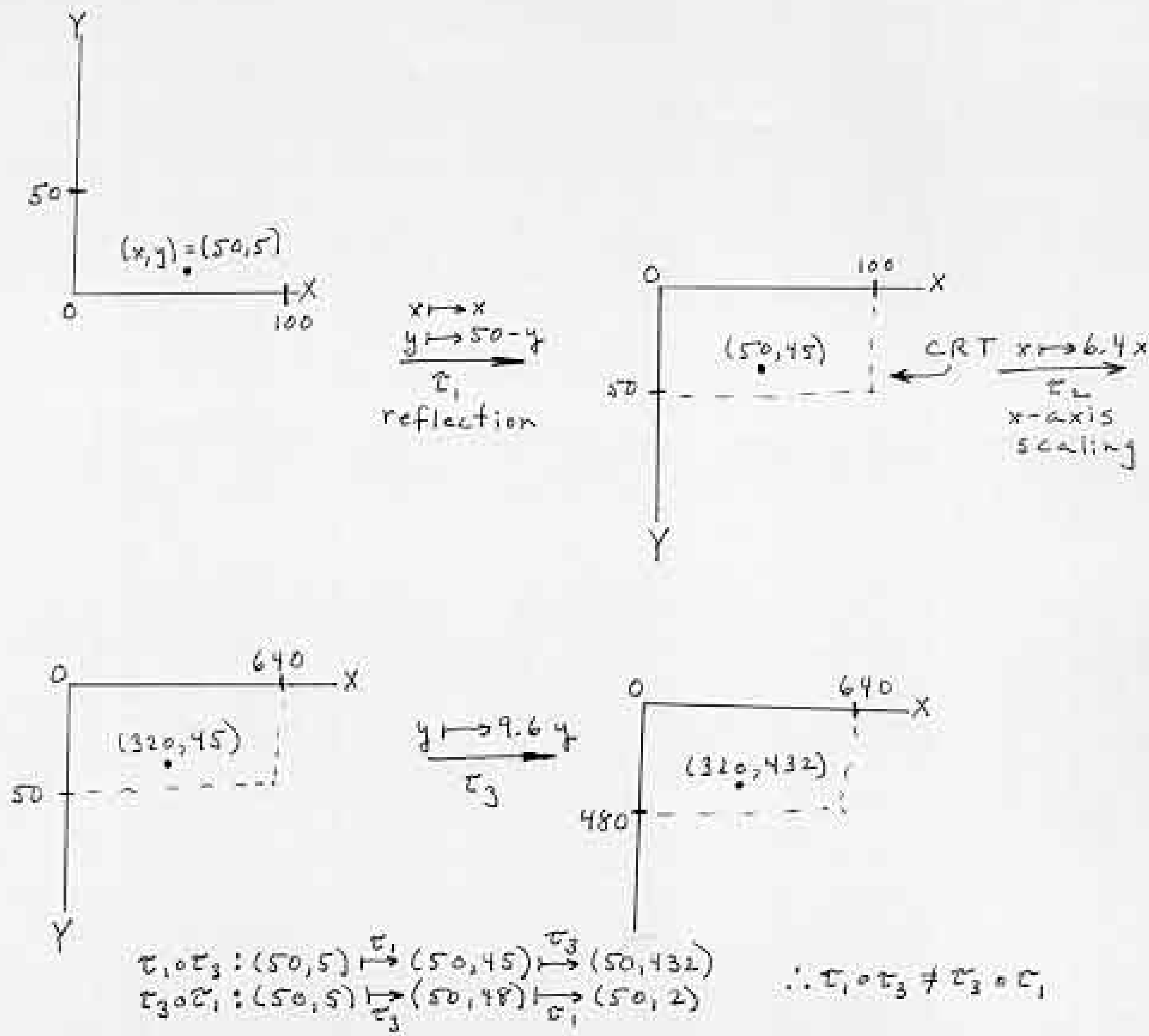


Figure VII.2

rubber sheet them to an irregularly spaced one).

v. Any ellipse is affine equivalent to a circle (demonstrated via copier technology).

a. Parallelism and GIS: crossing lines and polygon area.

Groups suggest how theoretical structure may be built from assembling simple pieces. GIS algorithms for complex processes are also often built from assembling simple pieces.

Straight lines

How can we tell if two lines intersect in a node?

Example from NCGIA Lecture 32: does the line L_1 from (4,2) to (2,0) cross the line L_2 from (0,4) to (4,0)? From a mathematical standpoint, two lines in the Euclidean plane cross if they have different slopes, m_1 and m_2, where the slope m between points (x_1, y_1) and (x_2, y_2) is calculated as $(y_2 - y_1)/(x_2 - x_1)$. In this case, the slope of L_1 is $(0-2)/(2-4) = 1$

and the slope of L_2 is $(0-4)/(4-0) = -1$. The slopes are different, so the lines cross in the plane. However, in the GIS context:

i. Do the lines cross on the computer screen, or is the intersection point outside the bounded Euclidean region of the screen?

ii. Even if the lines cross on the screen, do they intersect at a node of the data base (was that point digitized)?

To answer these questions, it is necessary to determine the intersection point of the two lines.

Equation of L_1: one form for the equation of a line between two points (x_1, y_1) and (x_2, y_2) is

$$y - y_1 = m(x - x_1)$$

where m is the slope and b is the second coordinate of the y-intercept. Thus, L_1 has equation $y - 2 = 1(x-4)$ or $y = x - 2$; L_2 has equation $y - 4 = -1(x-0)$ or $y = -x + 4$.

Solve these equations simultaneously to yield $x = 3$ and $y = 1$.

Thus, if the point $(3,1)$ lies within the boundaries of the screen, the lines intersect on the screen; if the point $(3,1)$ was digitized, then another line might be hooked onto the intersection point. If it was not digitized, then the lines "cross" but do not intersect, much as water pipes might cross but do not necessarily intersect (as in snapping a segment onto the middle of a line on the CRT). This is a graph-theoretic characteristic.

Note that vertical lines are a special case; their slope is undefined because $x_2 - x_1$, the denominator in the slope, is zero. Recognizing vertical lines should not be difficult, but it should be remembered that attempting to calculate slope across an entire set of lines, which might include vertical lines, can produce errors.

Chains of straight line segments.

How can we tell if chains of segments cross?

Because chains are of finite length and are bounded, it is possible to enclose them in a rectangle (no larger than the CRT screen) (Figure VII.3). This is a minimum enclosing rectangle.

Thus, given two chains, C_1 and C_2, if their respective minimum enclosing rectangles do not intersect (as do straight lines) then they do not intersect, and further testing is warranted.

Polygon area:

Calculate polygon area using notion of parallelism (Figure VII.4)

Simple rule, based on vertical lines, to determine if a point is inside or outside a polygon (Figure VII.5)

Centroids of polygons, with attached weights are often used as single values with which to characterize the entire polygon. Centroids are preserved, as centroids, under affine transformations.

These are technical procedures for determining various useful measures and are documented in NCGIA material; all are based in the theory of affine transformations applied to sets of pixels. Move now to consider the mechanics of how sets of affine transformations

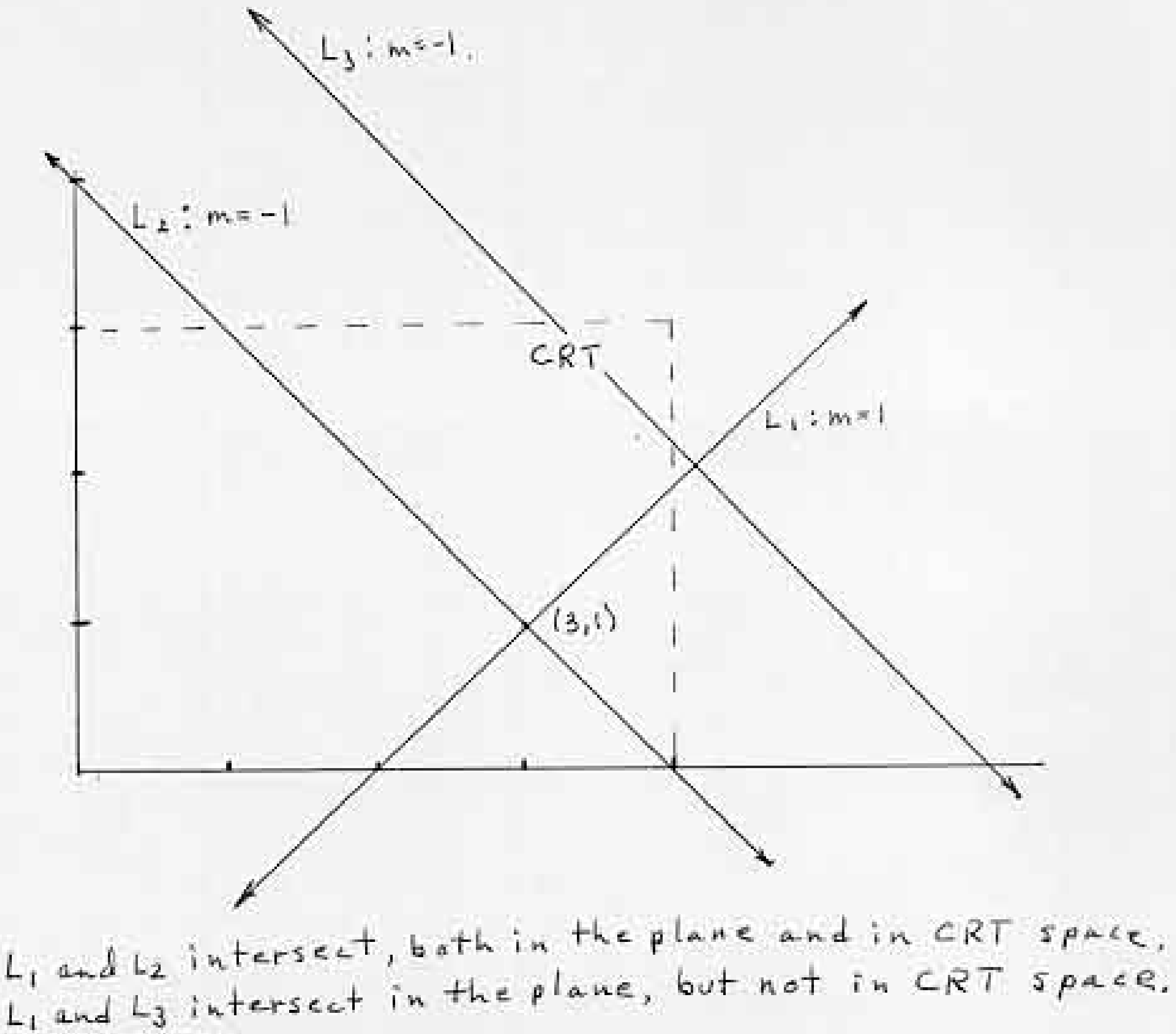


Figure VII.3

might affect a single pixel.

B. Group of symmetries of a square (pixel); the hexagonal pixel.

A square may have a set of rotations and of reflections applied to it as noted in Figure VII.6. Each may be represented as a permutation of the vertices, labelled clockwise. Permutations are multiplied as indicated in the example, below: multiply the permutation (1234) by the permutation (13)(24):

1 goes to 2 (in the left one)

and 2 goes to 4 (in the right one)

so 1 goes to 4 (in the product)

4 goes to 1 (in the left one)

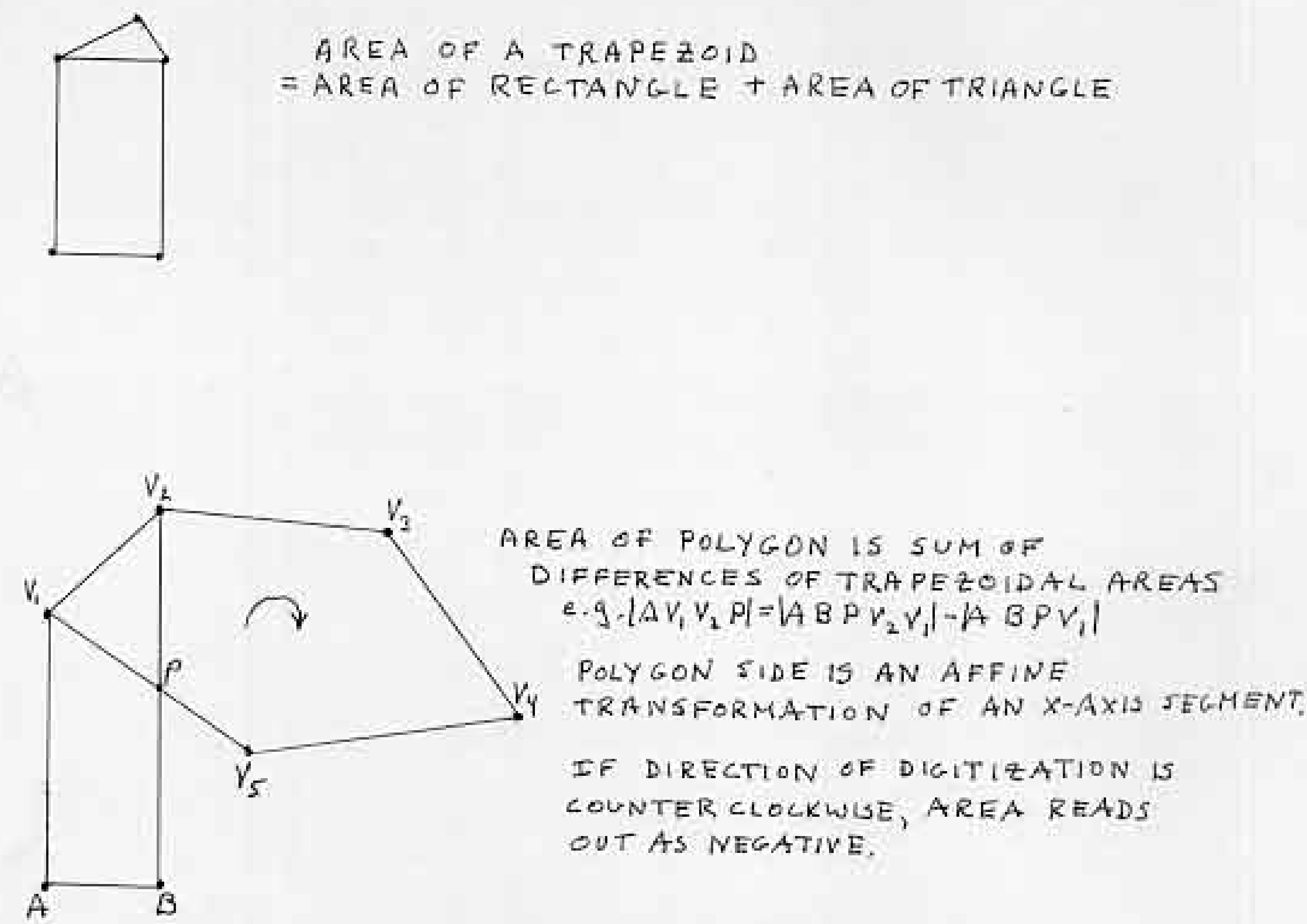


Figure VII.4

and 1 goes to 3 (in the right one)

so 4 goes to 3 (in the product)

3 goes to 4 (in the left one)

and 4 goes to 2 (in the right one)

so 3 goes to 2 (in the product)

2 goes to 3 (in the left one)

and 3 goes to 1 (in the right one)

so 2 goes to 1 (in the product)

This last stage is akin to snapping a polygon closed in a GIS environment—here it is a cycle of numbers rather than of vertices. Figure VII.6 shows all the calculations; note, that

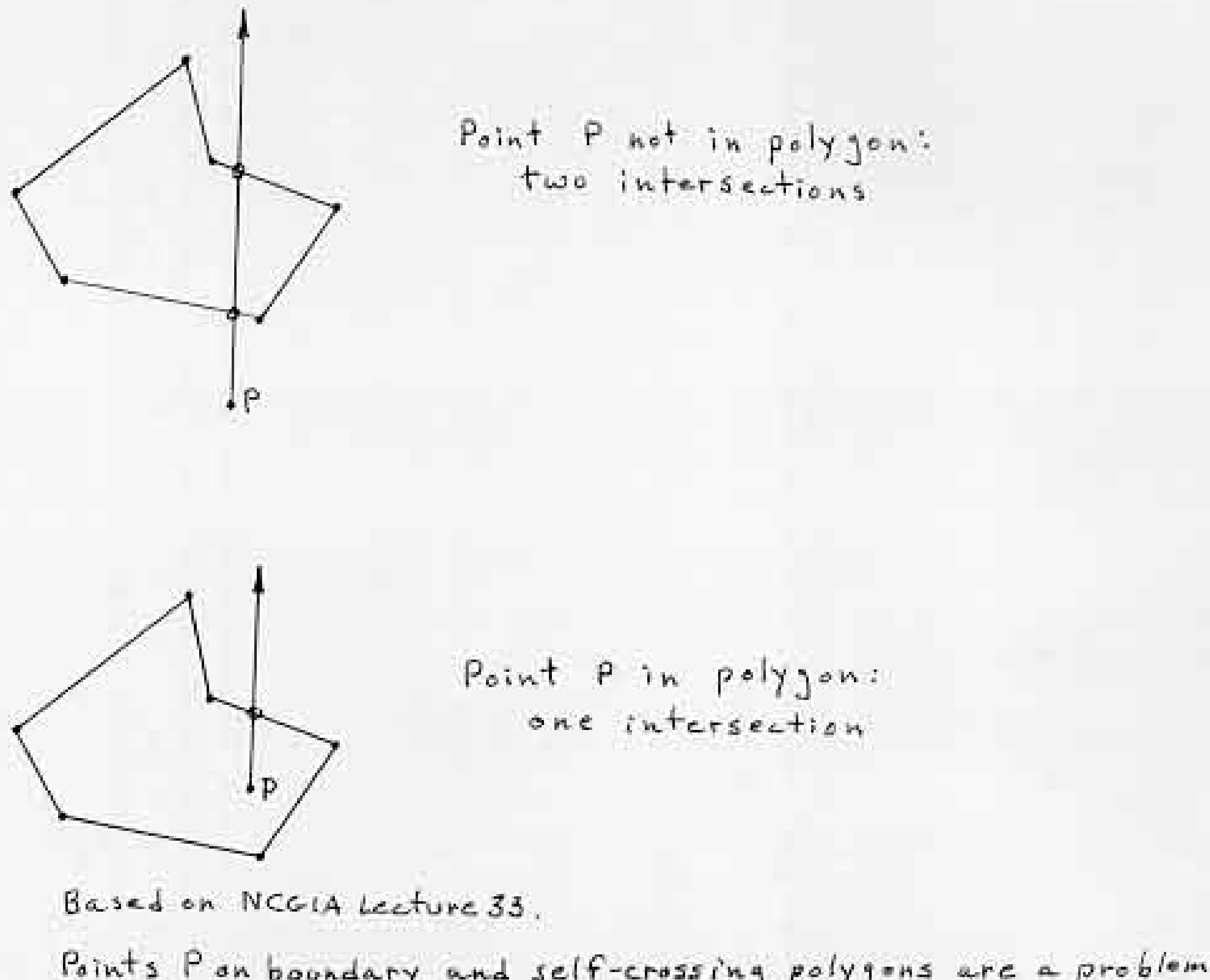


Figure VII.5

no new permutations ever arise; hence, the system is closed under $\star$; the rotation I serves as the identity transformation; each element has an inverse:

$$I \star I = I; I^{-1} = I$$

$$R_1 \star R_3 = I; R_1^{-1} = R_3$$

$$R_2 \star R_2 = I; R_2^{-1} = R_2$$

$$R_3 \star R_1 = I; R_3^{-1} = R_1$$

$$H \star H = I; V \star V = I; D_1 \star D_1 = I; D_2 \star D_2 = I.$$

So, this system is a "group." It is not, however, a commutative group—for example, $R_1 \star H = D_2$ and $H \star R_1 = D_1$. Once again, a reminder to be careful when combining reflections with affine transformations. Note that the set of rotations (including the identity rotation)

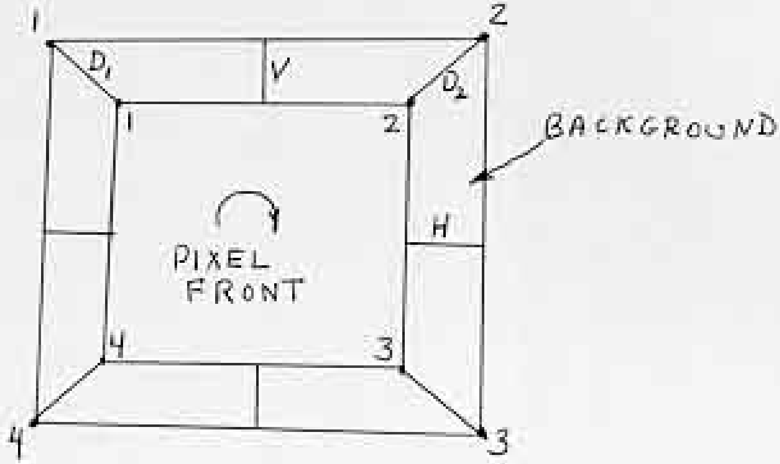


Figure VII.6 Group of symmetries of a square

Rotations:	Permutation representation
I: identity	(1)(2)(3)(4)
R_1: through 90 deg	(1234)
R_2: through 180 deg	(13)(24)
R_3: through 270 deg	(1432)
Reflections:	**Permutation representation**
H: horizontal	(14)(23)
V: vertical	(12)(34)
D_1: diagonal, 1 to 3	(1)(3)(24)
D_2: diagonal, 2 to 4	(2)(4)(13)

Table–operation, $\star$, is multiplication of permutations.

$\star$	I	R_1	R_2	R_3	H	V	D_1	D_2
I	I	R_1	R_2	R_3	H	V	D_1	D_2
R_1	R_1	R_2	R_3	I	D_2	D_1	H	V
R_2	R_2	R_3	I	R_1	V	H	D_2	D_1
R_3	R_3	I	R_1	R_2	D_1	D_2	V	H
H	H	D_1	V	D_2	I	R_2	R_1	R_3
V	V	D_2	H	D_1	R_2	I	R_3	R_1
D_1	D_1	V	D_2	H	R_3	R_1	I	R_2
D_2	D_2	H	D_1	V	R_1	R_3	R_2	I

is itself a group within this group. This is a "subgroup"—it is commutative—the order in which rotations are applied to the square is irrelevant.

Figure VII.6, continued

Permutation	⋆	Permutation	=	Permutation	
(1)(2)(3)(4)		(1)(2)(3)(4)		(1)(2)(3)(4)	I
(1234)		(1)(2)(3)(4)		(1234)	R_1
(13)(24)		(1)(2)(3)(4)		(13)(24)	R_2
(1432)		(1)(2)(3)(4)		(1432)	R_3
(14)(23)		(1)(2)(3)(4)		(14)(23)	H
(12)(34)		(1)(2)(3)(4)		(12)(34)	V
(1)(3)(24)		(1)(2)(3)(4)		(1)(3)(24)	D_1
(2)(4)(13)		(1)(2)(3)(4)		(2)(4)(13)	D_2
(1)(2)(3)(4)		(1234)		(1234)	R_1
(1234)		(1234)		(13)(24)	R_2
(13)(24)		(1234)		(1432)	R_3
(1432)		(1234)		(1)(2)(3)(4)	I
(14)(23)		(1234)		(1)(3)(24)	D_1
(12)(34)		(1234)		(2)(4)(13)	D_2
(1)(3)(24)		(1234)		(12)(34)	V
(2)(4)(13)		(1234)		(14)(23)	H
(1)(2)(3)(4)		(13)(24)		(13)(24)	R_2
(1234)		(13)(24)		(1432)	R_3
(13)(24)		(13)(24)		(1)(2)(3)(4)	I
(1432)		(13)(24)		(1234)	R_1
(14)(23)		(13)(24)		(12)(34)	V
(12)(34)		(13)(24)		(14)(23)	H
(1)(3)(24)		(13)(24)		(2)(4)(13)	D_2
(2)(4)(13)		(13)(24)		(1)(3)(24)	D_1
(1)(2)(3)(4)		(1432)		(1432)	R_3
(1234)		(1432)		(1)(2)(3)(4)	I
(13)(24)		(1432)		(1234)	R_1
(1432)		(1432)		(13)(24)	R_2
(14)(23)		(1432)		(12)(34)	D_2
(12)(34)		(1432)		(14)(23)	D_1
(1)(3)(24)		(1432)		(2)(4)(13)	H
(2)(4)(13)		(1432)		(1)(3)(24)	V
(1)(2)(3)(4)		(14)(23)		(14)(23)	H
(1234)		(14)(23)		(2)(4)(13)	D_2
(13)(24)		(14)(23)		(12)(34)	V
(1432)		(14)(23)		(1)(3)(24)	D_1
(14)(23)		(14)(23)		(1)(2)(3)(4)	I
(12)(34)		(14)(23)		(13)(24)	R_2
(1)(3)(24)		(14)(23)		(1432)	R_3
(2)(4)(13)		(14)(23)		(1234)	R_1
(1)(2)(3)(4)		(12)(34)		(12)(34)	V
(1234)		(12)(34)		(1)(3)(24)	D_1
(13)(24)		(12)(34)		(14)(23)	H

(1432)	(12)(34)	(2)(4)(13)	D_2
(14)(23)	(12)(34)	(13)(24)	R_2
(12)(34)	(12)(34)	(1)(2)(3)(4)	I
(1)(3)(24)	(12)(34)	(1234)	R_1
(2)(4)(13)	(12)(34)	(1432)	R_3
(1)(2)(3)(4)	(1)(3)(24)	(1)(3)(24)	D_1
(1234)	(1)(3)(24)	(14)(23)	H
(13)(24)	(1)(3)(24)	(2)(4)(13)	D_2
(1432)	(1)(3)(24)	(12)(34)	V
(14)(23)	(1)(3)(24)	(1234)	R_1
(12)(34)	(1)(3)(24)	(1432)	R_3
(1)(3)(24)	(1)(3)(24)	(1)(2)(3)(4)	I
(2)(4)(13)	(1)(3)(24)	(13)(24)	R_2
(1)(2)(3)(4)	(2)(4)(13)	(2)(4)(13)	D_2
(1234)	(2)(4)(13)	(12)(34)	V
(13)(24)	(2)(4)(13)	(1)(3)(24)	D_1
(1432)	(2)(4)(13)	(14)(23)	H
(14)(23)	(2)(4)(13)	(1432)	R_3
(12)(34)	(2)(4)(13)	(1234)	R_1
(1)(3)(24)	(2)(4)(13)	(13)(24)	R_2
(2)(4)(13)	(2)(4)(13)	(1)(2)(3)(4)	I

Are there any other subgroups? Yes, I, R_2, H, V also form a commutative subgroup. Note that the product of two reflections is a rotation.

A similar style of analysis might be executed for the pixel viewed as a hexagon. Other theoretical issues arise concerning the possibility of using a crt display with hexagonal pixels.

i. Issues involving centroids

a. Transformation to generate a centrally- symmetric hexagon from an arbitrary (convex) hexagon (rubbersheeting; TIN).

One such issue involves concern for taking a set of irregularly-spaced data points and converting them into some sort of more regular distribution (as with rubbersheeting and a TIN). This procedure illustrates how to transform an arbitrary convex hexagon (V_1, V_2, V_3, V_4, V_5, V_6) into a centrally symmetric hexagon (S_1, S_2, S_3, S_4, S_5, S_6) centered on a point that is easy to find. (See construction in *Solstice I*—Summer, 1990, Vol. I, No. 1., pp. 41-42.) Thus, rubbersheeting would appear possible with an hexagonal pixel.

b. Area algorithm generalizes to hexagons: regular hexagon is two isosceles trapezoids (one on either side of a single diameter of the hexagon).

What else might generalize from the square pixel format to the hexagonal pixel format? A hexagon can be decomposed into two trapezoids; thus one might imagine using an algorithm similar to that for the square pixel to find polygon areas relative to an hexagonal pixel display.

c. Steiner networks as boundaries of sets of hexagonal pixels; given a set of points, find a minimal hexagonal network linking them.

If centers of gravity (centroids) are used as a centering scheme in a triangulated irregular network (or other network of polygons), then it would be nice to have no centroid lie outside a triangular cell (or other polygon). A centroid is the intersection point of medians; it is the balance point on which the figure would rest. Sometimes the centroid lies outside the polygon; Coxeter suggests viewing the centroid as a balance point among electrical charges, thereby allowing for this possibility. Another point that is useful for using as a "central" weight is a Steiner point; in a triangle, it is that point which minimizes total network length joining the three vertices. It is always within the triangle when no angle of the triangle is greater than or equal to 120 degrees. (See *Solstice-I*, Vol. I., no. 2, "Super-definition resolution.")

Assigning point weights to represent polygon values is one way to compare them; another way is to assign centrally-located networks traversing underlying grid lines (Manhattan lines with square pixels, Steiner networks with hexagonal pixels); another way is to overlay the areas—again, a point-line-area classification as mentioned in detail in one of Nystuen's earlier lectures.

ii. Issues involving polygon overlays.

a. Close-packings of hexagons; central place geometry.

b. Fractal approach; space-filling; data compression.

Polygon overlay is familiar from OSUMAP. Look at some abstract geographic/geometric issues that might suggest directions to consider in looking at ideas behind the process of overlays.

Geometry of central place theory–including fractal generation of these layers. Look for a number of issues of this sort, that are theoretical, in using GIS-type equipment. Below is an outline of material in these lectures and of suggestions for future directions in which to look.

I. Introduction: the role of theory. Mathematics is fundamental, and in dealing with spatial phenomena, geometry in particular, is fundamental. Historical precedent from Biology in works of D'Arcy Thompson; Tobler's map transformations.

A. Statement of Thompson regarding the role of theory.

B. Visual evidence: one species of fish is transformed into another actual species by choosing a suitable coordinate transformation.

II. Transformations.

A. Well-defined (single-valued).

B. Reversible

i. One-to-one correspondence

ii. Transformations of X onto Y.

C. "Rubbersheeting"—example from Nystuen lecture, with fire stations. What is involved is creating a transformaton from an irregular scatter of locations to a regular one, locating new points (fire stations) and snapping the surface back to the irregular scatter. This requires transformations that are reversible.

III. Types of transformations and examples.

A. Affine

i. Translation

ii. Scaling

iii. Rotation

iv. Reflection

B. Curvilinear

IV. Exercise—scaling to make digitized map mesh with CRT scale.

V. GIS tie to Steiner networks.

VI. Digital topology. Quadtrees–Rosenfeld, Tobler. Jordan Curve Theorem; American Mathematical Society special sessions on digital topology (run by Rosenfeld). Hexagonal pixels—scanner technology.

VII. Local scale of mathematical extension of the concept of "affine transformation." The algebra of symmetry: definition of a group.

A. The affine group; affine geometry.

i. Parallelism and GIS: crossing lines and polygon area.

ii. Projective geometry; any two lines intersect in a point; no parallels. Here for completeness–not really discussed.

B. Group of symmetries of a square (pixel); the hexagonal pixel.

i. Issues involving centroids.

a. Transformation to generate a centrally-symmetric hexagon from an arbitrary (convex) hexagon (rubbersheeting; TIN).

b. Area algorithm generalizes to hexagons; hexagon is two trapezoids.

c. Steiner networks as boundaries of sets of hexagonal pixels; given a set of points, find a minimal hexagonal network linking them–dealt with in a third lecture, not presented here.

ii. Issues involving polygon overlays.

a. Close-packings of hexagons; central place geometry.

b. Fractal approach; space-filling; data compression.

VIII. Global scale of mathematical extension of the concept of "affine transformation." Topology.

A. Combinatorial topology.

i. Jordan curve theorem. GIS connection, inside and outside of polygons.

ii. Cell complexes; 0, 1, and 2 cells of GIS.

iii. Hexagons derived from barycentric subdivision of a complex.

B. Point-set topology.

i. Definitions.

ii. Consequences of Definitions interpreted in GIS context.

C. Digital topology.

i. Jordan curve theorem–3-dimensions.

ii. Quadtrees.

III. Further extension at different scales. Commutative diagrams—entry to different level of mathematical thought and spatial theory.

Publications of the Institute of Mathematical Geography have been reviewed in

1. *The Professional Geographer* published by the Association of American Geographers;
2. *The Urban Specialty Group Newsletter* of the Association of American Geographers;
3. *Mathematical Reviews* published by the American Mathematical Society;
4. *The American Mathematical Monthly* published by the Mathematical Association of America;
5. *Zentralblatt* Springer-Verlag, Berlin
6. *Mathematics Magazine*

SOLSTICE

Institute of Mathematical Geography

Journal of the
Institute of Mathematical Geography

Winter, 1991

The purpose of *Solstice* is to promote interaction between geography and mathematics. Articles in which elements of one discipline are used to shed light on the other are particularly sought. Also welcome, are original contributions that are purely geographical or purely mathematical. These may be prefaced (by editor or author) with commentary suggesting directions that might lead toward the desired interaction. Individuals wishing to submit articles, either short or full- length, as well as contributions for regular features, should send them, in triplicate, directly to the Editor-in-Chief. Contributed articles will be refereed by geographers and/or mathematicians. Invited articles will be screened by suitable members of the editorial board. IMaGe is open to having authors suggest, and furnish material for, new regular features.

Send all correspondence to:

Institute of Mathematical Geography

2790 Briarcliff

Ann Arbor, MI 48105-1429

(313) 761-1231

IMaGe@UMICHUM

This document is produced using the typesetting program, TEX, of Donald Knuth and the American Mathematical Society. Notation in the electronic file is in accordance with that of Knuth's *The TEXbook*. The program is downloaded for hard copy for on The University of Michigan's Xerox 9700 laser– printing Xerox machine, using IMaGe's commercial account with that University.

Unless otherwise noted, all regular features are written by the Editor–in–Chief.

Upon final acceptance, authors will work with IMaGe to get manuscripts into a format well–suited to the requirements of *Solstice*. Typically, this would mean that authors would submit a clean ASCII file of the manuscript, as well as hard copy, figures, and so forth (in camera–ready form). Depending on the nature of the document and on the changing technology used to produce *Solstice*, there may be other requirements as well. Currently, the text is typeset using TEX; in that way, mathematical formulæ can be transmitted as ASCII files and downloaded faithfully and printed out. The reader inexperienced in the use of TEX should note that this is not a "what–you–see–is–what–you–get" display; however, we hope that such readers find TEX easier to learn after exposure to *Solstice*'s e-files written using TEX!

Copyright will be taken out in the name of the Institute of Mathematical Geography, and authors are required to transfer copyright to IMaGe as a condition of publication. There are no page charges; authors will be given permission to make reprints from the electronic file, or to have IMaGe make a single master reprint for a nominal fee dependent on manuscript length. Hard copy of *Solstice* is available at a cost of $15.95 per year (plus shipping and handling; hard copy is issued once yearly, in the Monograph series of the Institute of Mathematical Geography. Order directly from IMaGe. It is the desire of IMaGe to offer electronic copies to interested parties for free. Whether or not it will be feasible to continue distributing complimentary electronic files remains to be seen. Presently *Solstice* is funded by IMaGe and by a generous donation of computer time from a member of the Editorial Board. Thank you for participating in this project focusing on environmentally-sensitive publishing.

ISBN: 1-877751-53-7 ISSN: 1059-5325

SUMMARY OF CONTENT

1. REPRINT.

Saunders Mac Lane. Proof, Truth, and Confusion. Given as the Nora and Edward Ryerson Lecture at The University of Chicago in 1982. Republished with permission of The University of Chicago and of the author.

I. The Fit of Ideas. II. Truth and Proof. III. Ideas and Theorems. IV. Sets and Functions. V. Confusion via Surveys. VI. Cost-benefit and Regression. VII. Projection, Extrapolation, and Risk. VIII. Fuzzy Sets and Fuzzy Thoughts IX. Compromise is Confusing.

2. ARTICLE

Robert F. Austin. Digital Maps and Data Bases: Aesthetics versus Accuracy *

I. Introduction. II. Basic Issues. III. Map Production. IV. Digital Maps. V. Computerized Data Bases. VI. User Community.

3. FEATURES

i. **Press Clipping** — reprint of "Briefing" from *Science* mentioning *Solstice* — "Online Journals," by Joseph Palca.

ii. **Educational feature** — Word search puzzle promoting mathematical and geographical awareness.

iii. **Software Briefs** — Brief descriptions of software provided by the creator. Look for reviews of the software in subsequent issues of *Solstice* .

a. RangeMapper TM — version 1.4. Created by Kenelm W. Philip, Tundra Vole Software, Fairbanks, Alaska. Program and Manual by Kenelm W. Philip; commentary from the Manual. "A utility for biological species range mapping, and similar mapping tasks in other fields."

b. "XYNIMAP" — created by David H. Douglas, University of Ottawa; "a comprehensive system for computer cartography and geo-spatial analysis." Preliminary Version.

iv. Index to Volumes I (1990) and II (1991) of *Solstice*.

v. Other publications of the Institute of Mathematical Geography.

vi. Solution to Word Search Puzzle.

ACKNOWLEDGMENTS

The 1982 Ryerson Lecture, "Proof, Truth, and Confusion," is republished in *Solstice* with the permission of The University of Chicago and with the permission of Professor Saunders Mac Lane.

The column "Online Journals" is republished in *Solstice* with the permission of the American Association for the Advancement of Science. It appeared originally in *Science*, "Briefings," Vol. 254, No. 5036 (Nov. 29, 1991), p. 1291. Copyright is held by the AAAS.

1. REPRINT

PROOF, TRUTH, AND CONFUSION

Saunders Mac Lane
Max Mason Distinguished Service Professor of Mathematics
The University of Chicago

The 1982 Nora and Edward Ryerson Lecture
at The University of Chicago

Saunders Mac Lane

Introduction
by
Hanna H. Gray
President of the University

Each year, one of the annual events at the University is the selection of the Nora and Edward Ryerson Lecturer. The selection is made by a committee of faculty which receives nominations from their faculty colleagues. Each year, this committee comes up with an absolutely superb selection for the Ryerson Lecturer, and this year is triumphant confirmation of that generalization.

The selection emanates from faculty nomination and discussion, and it is analogous to the process of the selection of faculty in this University, representing a selection based on the work and contribution, on the high esteem for the intellectual imagination and breadth of a colleague. The peer review process, in this as in faculty appointments, stresses scholarship and research, stresses the contribution of a member of the faculty to the progress of knowledge. In addition, of course, the faculty appointment process looks also to teaching and to institutional citizenship.

If I had the nerve to fill out an A-21 form for Professor Saunders Mac Lane, I would, I think, be creating a new mathematics because I would award him 100 percent for research, 100 percent for teaching, and 100 percent for contribution or citizenship. Even I can add that up to 300 percent.

Now, of course, in the evaluation of younger scholars for junior appointments similar judgments are made. They are based on the same three categories, and they are judgments about the promise of continuing creativity, continuing growth, continuing intellectual contribution. That judgment of the young Saunders Mac Lane was made a very long time ago in Montclair, New Jersey. Montclair, New Jersey is the home of the Yale Club of Montclair. I once had the enormous privilege of being invited to the Yale Club of Montclair where I was given something called the Yale Bowl, which had on it an inscription testifying that I had earned my "Y" in the "Big Game of Life."

In 1929, the young Saunders Mac Lane went to the Yale Club of Montclair, I was told—he was then finishing his senior year at Yale—and there was a young dean of the Yale Law school who was leaving in order to go to the University of Chicago as president. And the Yale Club of Montclair, which usually gave its awards to football players, decided on this occasion to give recognition to a young mathematician and to a young law school dean, and that was where Mr. Hutchins and Mr. Mac Lane met.

As Saunders Mac Lane graduated from Yale, Mr. Hutchins encouraged him personally to come to Chicago. And Saunders came. He had, however, neglected to take steps that are usually taken when one travels to enter another university, and the chairman of the Department of Mathematics, Mr. Bliss, had to say to him rather directly, "Young man, you've got to apply first."

He did apply, and fortunately he was accepted. Within a year, he had received his M.A. from Chicago and had come into contact with lots of extraordinary people, but two very extraordinary people in particular. One was the great mathematician E. H. Moore, and the other was a graduate student in economics named Dorothy Jones, who was in 1933 to become Mrs. Mac Lane.

Now, those of us who know Saunders think of him as a Hyde Parker, and indeed as a Hyde Parker forever. And, of course, he is a Hyde Parker, and a Hyde Parker forever, but he had a period in his life, I have to tell you, after he had taken his M.A., when he became something of an academic traveler. We really ought to have been able to trace those travels when we think about Saunders' choice of costume.

Now, that's not easy to figure out today because I think that necktie came out of a safe this morning. But if you think about the flaming reds, for example, that Saunders affects, you are perhaps reminded of Cambridge, Massachusetts. If you think of the Scottish plaids which he affects, that's a harder one, because I would say that that has to do with the great tradition which took him to New England and made him for a time a resident of Connecticut. And then, of course, there is the Alpine hat which could only have come from Ithaca, New York.

Saunders received his doctorate from the University of Göttingen in 1934. He had spent the years 1933-34 again at Yale as a Sterling Fellow. He then spent two years at Harvard. He then spent a year at Cornell. Then he came to the University of Chicago for a year. And then again he moved, called back to Harvard as an assistant professor, and there he rapidly went through the ranks. Fortunately, in 1947, he returned to the University of Chicago and, in 1963, became the Max Mason Distinguished Service Professor. Between 1952 and 1958, he succeeded Marshall Stone as chairman of the Department of Mathematics for two three-year terms, and he has served the University as he has his department with total dedication.

Saunders has extended his role beyond our University, serving primarily and prominently in a number of national scholarly organizations and institutions devoted to large questions of the relationship of learning to policy. He was president of the Mathematical Association of America and received its Distinguished Service Award in 1975 in recognition of his sustained and active concern for the advancement of undergraduate mathematical teaching and undergraduate mathematics. He was also president of the American Mathematical Society in 1973-74. He has been a member of the National Science Board and vice-president of the National Academy of Sciences.

His work in mathematics, of course, has been widely recognized. Alfred Putnam, who studied with Saunders at Harvard, had this to say of Saunders in a biographical sketch that he has published. He wrote, "Beginning as a graduate student with a brief exposure to group extensions, I've watched the development of Saunders Mac Lane's mathematics through homological algebra to category theory. Saunders Mac Lane belongs in a category by himself."

And so he does. So he does as a mathematican, as an academic citizen, as a spokesman for the fundamental values and principles of the University, and, of course, in sartorial wonder.

Now, it is to this category that we look for the Nora and Edward Ryerson Lecturers. When the Trustees established the lectureship in 1973, they sought a way to celebrate the relationship that the Ryersons and their family have had with our University—a relationship of shared values and a commitment to learning at the most advanced level.

Mr. Ryerson was elected to the Board in 1923 and became Chairman of the Board in 1953. Nora Butler Ryerson was a founding member, if not *the* founder, of the University's Women's Board. Both embraced a civic trust that left few institutions in our city untouched, and they passed to future generations of their family the sense of engagement and participation.

Saunders Mac Lane, through his staunch loyalty to our University, his broad interest in the community of scholars and their work, his distinguished scholarly career, represents these values for us in a special way, and, of course, he is entirely uncompromising also in his commitment to them. It is a pleasure to introduce this year's Ryerson Lecturer, Saunders Mac Lane.

Proof, Truth, and Confusion

Saunders Mac Lane

I. The Fit of Ideas

It is an honor for a mathematician to stand here. Let me first say how much I appreciate the initiative taken by the trustees on behalf of the Ryerson family in providing for this series of lectures, which afford opportunity for a few fortunate faculty members to present aspects of their scholarly work which might be of interest to the whole university community. In my own case, though the detailed development of mathematics tends to be highly technical, I find that there are some underlying notions from mathematics and its usage which can and will be of general interest. I will try to disentangle these and to relate them to the general interest.

This intent accounts for my title. Mathematicians are concerned to find truth, or, more modestly, to find a few new truths. In reality, the best that I and my colleagues in mathematics can do is to find proofs which perhaps establish some truths. We try to find the right proofs. However, some of these proofs and the techniques and numbers which embody them have turned out to be so popular that they are applied where they do not belong—with results which produce confusion. For this, I will try to cite examples and to draw conclusions.

This involves a thesis as to the nature of mathematics: I contend that this venerable subject is one which does reach for truth, but by way of proof, and does get proof, by way of the concatenation of the right ideas. The ideas which are involved in mathematics are those ideas which are formal or can be formalized. However, they are not purely formal; they arise from aspects of human activity or from problems arising in the advance of scientific knowledge. The ideas of mathematics may not always lead to truth; for this reason it is important that good ideas not be confused by needless compromise. In brief, the ideas which matter are the ideas that fit.

However, the fit may be problematical. A friend of mine with a vacation home in Vermont wanted to suitably decorate his barn, and so asked the local painter to put on the door "the biggest number which can be written on the broad side of a barn door." The painter complied, painting on the barn door a digit 9 followed by as many further such digits as could be squeezed onto the door (Figure 1.a). A competitor then claimed he could do better by painting smaller 9's and so a bigger number. A second competitor then rubbed out the first line and wrote instead: The square of the number 9, ... (Figure 1.b). Even that didn't last, because another young fellow proposed the paradoxical words, "One plus the biggest number that can be written on the broad side of this barn door" (Figure 1.c). At each moment, this produces a bigger number than anything before. We may conclude that there is no such biggest number. This may illustrate the point that it is not easy to get the ideas that fit—on barn doors or otherwise.

II. Truth and Proof

I return to the "truth" of my title. When I was young I believed in RMH-which sometimes stands for Robert Maynard Hutchins, who to my great profit first encouraged me to come to Chicago—and which sometimes stands for the slogan, "Reach Much Higher." At any rate, when young I thought that mathematics could reach very much higher so as to achieve absolute truth. At that time, *Principia Mathematica* by Whitehead and Russell seemed to model this reach; it claimed to provide all of mathematics firmly founded on the

999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999

The biggest number on that barn door.

The Square of
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999
999,999,999,999,999

A bigger number on that barn door.

One Plus (+) the
Biggest Number
that can be
Written
on the
Broad Side
of this
Barn Door

An even bigger number on that barn door.

Figure 1. a. The biggest number on that barn door. b. A bigger number on that barn door. c. An even bigger number on that barn door.

truths of logic. The logic in *Principia* was elaborate, symbolic, and hard to follow. As a result, it took me some years to discover that *Principia Mathematica* was not a *Practica Mathematica*—much of mathematics, in particular most of geometry, simply wasn't there in *Principia*. For that matter, what was there didn't come exclusively from logic. Logic could provide a framework and a symbolism for mathematics, but it could not provide guidelines for a direction in which to develop.

This limitation was a shocking discovery. Logic, even the best symbolic logic, did not provide all of absolute truth. What did it provide instead? It provided proof—the rigorous proof of one formal statement from another prior statement; that is, the deduction of theorems from axioms. For such a deduction, one needed logic to provide the rules of inference. In addition, one needed the subject matter handled in the deductions: the ideas used in the formulation of the axioms of geometry and number theory, as well as the suggestions from outside mathematics as to what theorems might usefully be proved from these axioms.

Deductive logic is important not because it can produce absolute truth but because it can settle controversy. It has settled many. One notable example arose in topology, a branch of mathematics which studies qualitative properties of geometric objects such as spheres. From this perspective, a smooth sphere and a crinkly sphere would have the same qualitative properties—and we would consider not just the ordinary spheres—two-dimensional, since the surface has two dimensions—but also the spheres of dimensions 3, 4, and higher (Figure 2). For these spheres, topologists wished to calculate a certain number which measures the connectivity—a measure "two dimensions up" from the dimension of the sphere. The Soviet topologist L. Pontrjagin in 1938 stated that this desired measure was one. Others thought instead that the measure was two. In a related connection, the American reviewer of another paper by Pontrjagin wrote, "Both theorems (of Pontrjagin) contradict a previous statement of the reviewer. It is not easy to see who is wrong here." Fortunately, it was possible to see. With careful analysis of the proof, Pontrjagin did see who was wrong—and in 1950 published a statement correcting his 1938 error: that the measure of connectivity two dimensions up is not one, but two.

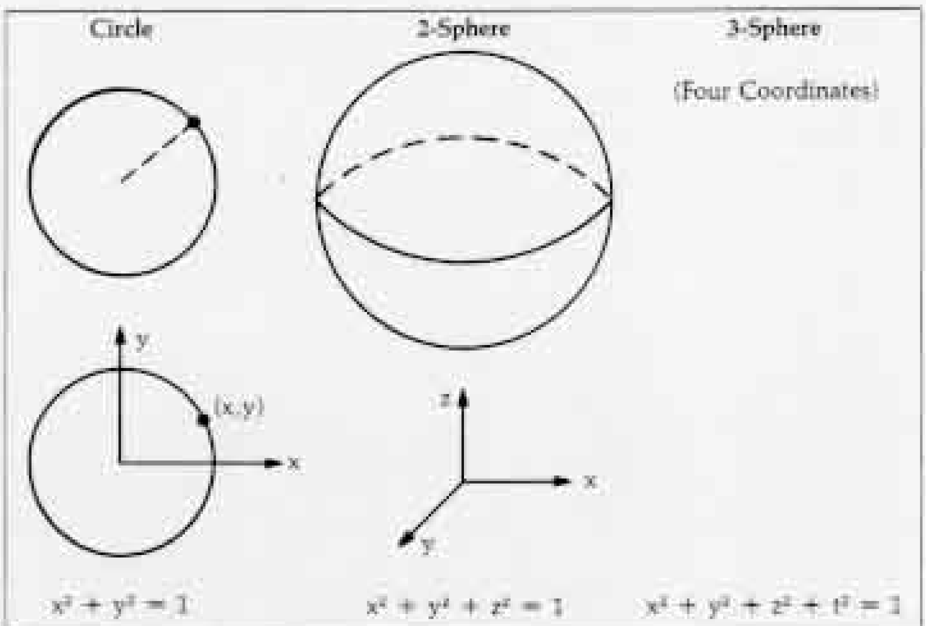


Figure 2

A few years ago, the *New York Times* carried an item about a similar fundamental disagreement between a Japanese topologist and one of our own recent graduate students, Raphael Zahler. Analysis of the deductions showed that Zahler was right. There lies the real role of logic: it provides a formal canon designed to disentangle such controversies.

Truth may be difficult to capture, but proof can be described with complete accuracy. Each mathematical statement can be written as a word or sentence in a fixed alphabet—using one letter for each primitive mathematical notion and one letter for each logical connective. A proof of a theorem is a sequence of such statements. The initial statement must be one of the axioms. Each subsequent statement not an axiom must be a consequence of prior statements in the sequence. Here "consequence" means "consequence according to one of the specified rules of inference"—rules specified in advance. A typical such rule is that of *modus ponens*: Given statments "S" and "S implies T," one may infer the statement "T."

This description gives a firm standard of proof. Actual proofs may cut a few corners or leave out some obvious steps, to be filled in if and when needed. Actual proofs may even be wrong. However, the formal description of a proof is complete and definitive. It provides a formal standard of rigor, not necessarily for absolute truth, but for absolute proof.

There is a surprising consequence: no one formal system suffices to establish all of mathematics. Precisely because there is such a rigorous description of a "proof" in a "formal system," Kurt Gödel was able to show that, in each such system with calculable rules of inference, one could formulate in the system a sentence which was not decidable in the system—that is, a sentence G which can neither be proved nor disproved according to the specified rules of inference. More exactly, this is the case for any system which contains the numbers and the rules of arithmetic, and in which the rules of inference can be explicitly listed or numbered in the fashion called "recursive."

In such a system, all statements are formal and are constructed from a fixed alphabet. Hence we can number *all* the possible proofs. Moreover, we can formulate within the system a sentence which reads, "n is the number of the proof of the statement with the number k." On this basis, and adapting ideas illustrated by the paradox of the barn door, one then constructs another sentence $G = G(p)$ (with number p) which reads, "There is no number which is the proof of the sentence number p." This means in particular that this very sentence G cannot be proved in the system. This is because G itself states that "there is no proof in the system for me"—hence G is true (Figure 3). Hence, unless the system is inconsistent, it can contain no refutation of G. Thus in such a formal system we can write one statement (and hence many) which, though true, is simply undecidable, yes or no, within the system.

Godel's Sentence G = G(p)

w = A chosen variable

Sub (x,t) = The number of the following statement
"In Statement no. x, Replace the Variable w by the Numeral t"

Dem (y, u) = "The Proof numbered y Demonstrates the Statement number u"

G(w) = "For no. y, Dem (y, Sub (w, w))

"Its number is p"

so Sub (p, p) is G(p)

G(p) = "There is no proof of G(p)"

"You can't prove me"

Figure 3

This result is startling. It may seem catastrophic—but it turns out to be not quite so disastrous. It shows that there is an intrinsic limitation on what can be proved within any *one formal* system; thus proof within one such system cannot give all of truth. Very well then, as we shall see, there can be more than one formal system and hence more than one way in which to reach by proof for the truth.

III. Ideas and Theorems

Some observers have claimed that mathematics is just formalism. They are wrong. A mathematical proof in a given formal system must be *about* something, but it is not about the outside world. I say it is about ideas. Thus the formal system of Euclidean geometry is about certain "pictorial" ideas: point, line, triangle, and congruence; in their turn, these ideas arose as means of formulating our spatial experiences of shape, size, and extent and our attempts to analyze motion and symmetry.

Each branch of formal mathematics has a comparable origin in some human activities or in some branch of scientific knowledge. In each such case, the formal mathematical system can be understood as the realization of a few central ideas.

Mathematics is built upon a considerable variety of such ideas—in the calculus, ideas about rate of change, summation, and limit; in geometry, ideas of proximity, smoothness, and curvature. To further illustrate what I mean here by "idea," I choose a small sample: The related ideas of "connect," "compose," and "compare."

To "Connect" means to join. There are different ways in which mathematicians have defined what it means for a piece of space to be connected. One definition says that a piece of space is connected if it does not fall apart into two (or more) suitably disjoint pieces. Another definition says that a piece of space is *path-connected* if any two points in the piece can be joined *within* the piece by a path—that is, by a continuous curve lying wholly in the piece. These two formal explications of the idea of "connected" are not identical; a piece of space which is path-connected is always connected in the first sense, but not necessarily vice-versa. This simple case of divergence illustrates the observation that the same underlying pre-formal idea can have different formalizations.

"Compose" is the next idea. To compose two numbers x and y by addition is to take their sum $x+y$; to compose them by multiplication is to take their product xy. To compose one motion L with a second motion M is to follow L by M to get the "composite" motion which we write as $L \circ M$. Thus to rotate a wheel first by 25° and then by 45° will yield after composition a rotation by 70°. To compose a path L connecting a point p to a point q with a path M connecting q to a third point s is to form the longer path $L \circ M$ which follows first L and then M, as in the top of figure 4. In all such cases of composition, the result of a composition $L \circ M \circ N$ of three things in succession depends on the factors composed and the sequence or order in which they were taken—but *not* on the position of the parenthesis. Thus arises one of the formal laws of composition, the associative law:

$$L \circ (M \circ N) = (L \circ M) \circ N.$$

However, $L \circ M$ may very well differ from $M \circ L$! The order matters.

The third sample idea is "Compare." One may compare one triangle with another as to size, so as to study congruent triangles. One may compare one triangle with another as to shape, and so study more generally similar triangles. Another comparison is that by

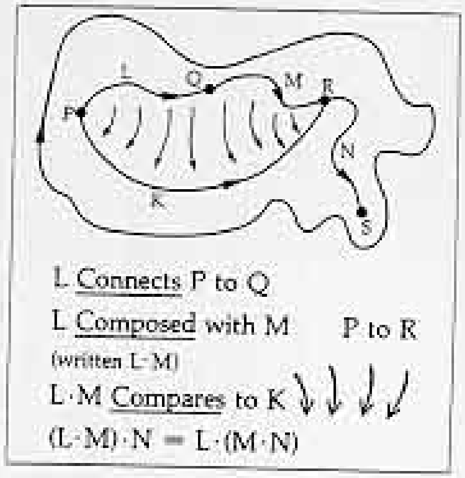


Figure 4.

deformation: Two paths in a piece of space may be compared by trying to deform the first path in a continuous way into the second—as in Figure 4, the composite path $L \circ M$ from p to r can be deformed smoothly and continuously into the path K also joining p to r.

Ideas such as these will function effectively in mathematics only after they have been formalized, because then explicit theorems about the ideas can be proved. The idea of composition is formalized by the concept of a group, which applies to those compositions in which each thing L being composed has an "inverse" thing or operation L^{-1} so that $L \circ L^{-1} = 1$. One readily sets down axioms for a group of "things" with such composition. The axioms are quite simple, but the concept has proven to be extraordinarily fruitful. There are very many examples of groups: Groups of rotations, groups of symmetry, crystallographic groups, groups permuting the roots of equations, the gauge groups of physics, and many others. There is a sense (analyzed by Eilenberg-Mac Lane in a series of papers) in which any group can be built up by successive extensions from certain basic pieces, called the "simple" groups. Specifically, a group is said to be *simple* when it cannot be collapsed into a smaller group except in a trivial way. A long-standing conjecture suggested that the number of elements in a finite simple group was necessarily either an even number or a prime number. About twenty years ago, here at Chicago, Thompson and Feit succeeded in proving this to be true (and I could take pleasure in the fact that Thompson, one of my students, had achieved such a penetrating result). The Thompson-Feit method turned out to be so suggestive and powerful that others have now been able to go on to explicitly determine *all* the finite simple groups. For example, the biggest sporadic one has $2^{46} \cdot 3^{20} \cdot 5^{9} \cdot 7^{6} \cdot 11^{2} \cdot 13^{3} \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 41 \cdot 47 \cdot 59 \cdot 71$ elements (that number is approximately 8 followed by 53 zeros). This simple group is called the "Monster" (Figure 5). Another one of our former students has been able to make a high dimensional geometric picture which shows that this monster really exists. He needed a space of dimension 196,884.

Groups also serve to measure the connectivity of spaces. In particular, there are certain homology groups which count the presence of higher dimensional holes in space. To start with, a piece χ of space is said to be *simply connected* if any closed path in the space χ

A Group G
Multiplication!

Any two elements g, h
have a Product g·h
Each g has an Inverse g^{-1}

All! Finite "Simple" Groups
One for Each Prime 2, 3, 5...
Several for each Dimension

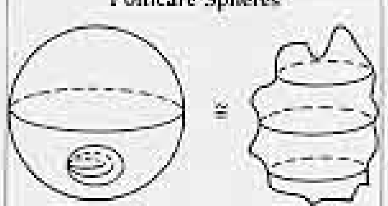
Plus "Sporadics"
For example, The Monster has

$2^{46} \cdot 3^{20} \cdot 5^{9} \cdot 7^{6} \cdot 11^{2} \cdot 13^{3} \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot$
$41 \cdot 47 \cdot 59 \cdot 71 \sim 8 \cdot 10^{53}$ elements

Figure 5

can be deformed into a point. For example, the surface of a sphere is simply connected, so its first homology group is zero; however, it has a non-zero second homology group—meaning the "hole" represented by the inside of the sphere. These properties characterize the two-dimensional sphere. Long ago, the French mathematician Poincaré said that the same should hold for a three-dimenional sphere (Figure 6). This famous conjecture has not yet been settled—but some years ago, Smale showed that the characterization was true for a sphere of dimension 5 or higher. Just during the last year, the Californian Michael Friedman, in a long proof, showed that it is also true for a sphere of dimension 4. Except for a solution which was announced on April 1, nobody yet knows the answer for a three-dimensional sphere. Proof advances, but slowly.

Poincaré Spheres



S^2 $x^2 + y^2 + z^2 = 1$ Connectivity:

Dimension	0	1	Points connect
	1	1	Circles collapse
	2	2	S^2 won't collapse

S^3 $x^2 + y^2 + z^2 + t^2 = 1$

Dimension	0	1	Points connect
	1	1	Circles collapse
	2	1	Spheres collapse
	3	2	S^3 won't collapse

Figure 6

IV. Sets and Functions

As already indicated, Whitehead and Russell, by *Principia Mathematica*, had suggested that all mathematical truth could be subsumed in one monster formal system. Their system, corrupted as it was with "types," was too complicated— but others proposed a system based on the idea of a *set*. A set is just a collection of things—nothing more. Mathematics does involve sets, such as the set of all prime numbers or the set of all rational numbers between 1 and 2. Mathematical objects can be defined in terms of sets. For example, a circle is the set of all points in the plane at a fixed distance from the center, while a line can be described as the set of all its points. Numbers can be defined as sets—the number two is the set of all pairs; an irrational number is the set of all smaller rational numbers. In this way numbers, spatial figures, and everything else mathematical can be defined in terms of sets (Figure 7). All that matters about a set S is the list of those things x which are members of S. When this is so, we write $x \in S$, and call this the "membership relation."

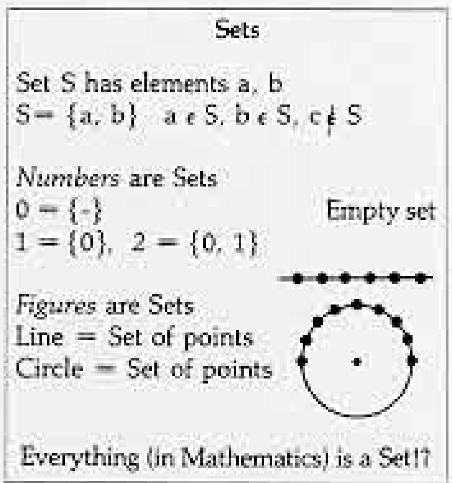


Figure 7

There are axioms (due to Zermelo and Fraenkel) which adequately formalize the properties of this membership relation. These axioms claim to provide a formal foundation–I call this the grand set-theoretic doctrine—for all of mathematics.

By 1940 or so this grand set-theoretic foundation had become so prominent in advanced mathematics that it was courageously taught to freshmen right here in the Hutchins college. This teaching practice spread nationally to become the keystone of the "New Math." As a result, twenty years later sets came to be taught in the kindergarten. There is even that story about the fond parents inquiring as to little Johnny's progress. Yes said the teacher, he is doing well in math except that he can't manage to write the symbol ϵ when x is a member of the set S.

Johnny was not the only one in trouble. The grand doctrine of the new math: "Everything is a set" came at the cost of making artificial and clumsy definitions. Moreover, putting everything in one formal system of axioms for set theory ran squarely into the difficulties presented by Gödel's undecidable propositions.

Fortunately, just about the time when sets reached down to the kindergarten, an alternative approach to a system of "all" (better "most") of mathematics turned up. This used again the idea of composition for functions $f : S \longrightarrow T$ sending the elements of a set S to some of those of another set T. Another function $g : T \longrightarrow U$ can then be composed with f to give a new function $g \circ f$ (Figure 8). It sends an element of S first by f into T and then by g into U. The prevalence of many such compositions led Eilenberg and Mac Lane in 1945 to define the formal axioms for such composition. With no apologies to Aristotle, they called such a system a "category"—because many types of mathematical objects did form such categories, and these properties were useful in the organization of mathematics. Note especially that the intuitive idea of "composition" has several different formalizations: category and group.

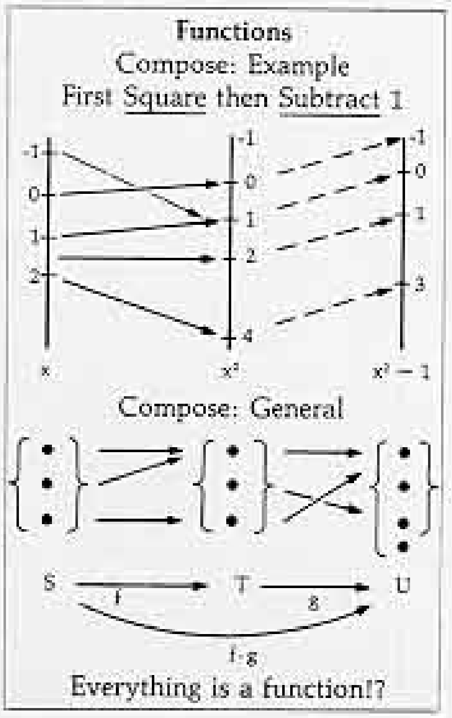


Figure 8

Then in 1970 Lawvere and Tierney made a surprising discovery: that in treating a function $f : S \longrightarrow T$ one could forget all about the elements in S and T, and write enough axioms on composition alone to do almost everything otherwise done with sets and elements. This formal system is called an "elementary topos"—to suggest some of its connections to geometry and "Top"-ology. Their success in discovering this wholly new view of mathematics emphasizes my fundamental observation: That the ideas of mathematics are various and can be encapsulated in different formal systems.

Waiting to be developed, there must be still other formal systems for the foundation and organization of mathematics.

V. Confusion via Surveys

The crux of any search for the right alternative to set theory is the search for the right concatenation of ideas—in the same way in which leading ideas in mathematics have been combined in the past to solve problems (the Poincaré conjecture on spheres) and to give new insights. Thus it was with our example, where the related ideas of connection, composition, and comparison came together in group theory, in the application of groups to geometry, and in category theory.

But sometimes the wrong ideas are brought together, or the right ideas are used in the wrong way. Today the use of numbers and of quantitative methods is so pervasive that many arrays of numbers and of other mathematical techniques are deployed in ways which do not fit.

This I will illustrate by some examples. Recently, in connection with my membership on the National Science Board, I came across the work of one prominent social scientist who was promoting (perhaps with reason) the use of computer-aided instruction in courses for college students. However, the vehicle he chose for such instruction was the formal manipulation of the elementary consequences of the Zermelo-Fraenkel axioms for set theory—and the result was an emphasis on superficial formalism with no attention to ideas or meaning. It was, in short, computer-aided pedantry.

"Opinion Surveys" provide another example of the confusion of ideas. For some social and behavioral research, the necessary data can be obtained only by survey methods, and responsible scientists have developed careful techniques to help formulate the survey questions used to probe for facts. Unfortunately these techniques are often used carelessly—both because of commercial abuse, statistical malpractice, or poor formulation of survey questions. First the malpractice:

On many surveys the percentage of response is uncomfortably low, with the result that the data acquired are incomplete. This situation has led the statisticains into very elaborate studies of means for approximately completing such "incomplete data." One recent and extensive such publication (by the National Research Council) seemed to me technically correct but very elaborate—perhaps overdone, and in any event, open to the misuse of too much massaging of data that are fatally incomplete.

In opinion surveys touching directly on the academic profession some of the worst excesses are those exhibited by the so-called "Survey of the American Professoriate." Successive versions of this survey are replete with tendentious and misleading questions, often such likely to "create" opinion rather than to measure actual existing opinions. Despite heroic attempts by others to suggest improvements, the authors of this particular survey have continued in their mistaken practices in new such surveys—as has been set forth with righteous indignation by Serge Lang in his publication *The File: A Case Study in Correction.*

That otherwise useful publication *Science Indicators* from the National Science Board makes excessive use of opinion surveys. The most recent report of the series (*Science Indicators 1980*) coupled results from a new, more carefully constructed opinion survey with a simple continuation of poorly formulated questions taken from previous and less careful surveys.

The main new opinion survey commissioned for this NSB report used an elaborate design—but this design still involved some basic misconceptions about science and some questions about science so formulated as to distort the opinions which were to be surveyed. For example, its Question 71 first observes that "Science and technology can be directed toward solving problems in many different areas"—while I would claim that science cannot be "directed" in the fashion intended by government bureaucrats. The question then lists fourteen areas and asks "Which three areas on the list would you *most* like to receive science and technology funding from your tax money?" Of the fourteen areas, some had little to do with science or technology and much to do with the political and economic structure of society (for example, controlling pollution, reducing crime, and conserving energy). Only one of the fourteen dealt with basic knowledge. With an unbalanced list of questions like this, the report goes on to claim that the answers "suggest that the public interest tends to focus on the practical and immediate rather than on results that are remote from daily life." This may be so, but it cannot be demonstrated by answers to a survey questionnaire which itself is so constructed as to focus on the "practical and immediate."

To get comparisons of opinions across time, new surveys try to continue questions which have been used before—and so often use older questions of a clearly misleading character. In *Science Indicators*, a typical such previous question is the hopelessly general one, "Do you feel that science and technology have changed life for the better or the worse?" The current version of this question does still more to lead the respondent to a negative answer. It reads, "Is future scientific research more likely to cause problems than to find solutions to our problems?" It is no wonder that this latter slanted question, in the 1979 survey, had only 60% answers favorable to science, while the earlier one had 75% favorable in 1974 and 71% in 1976.

Surveys also may pose questions which the respondents are in no position to answer. For instance, one question in this survey probed the respondents' expectations of scientific and technological achievements: "During the next 25 years or so, would you say it is very likely, possible but not too likely, or not likely at all that researchers will discover a way to predict when and where earthquakes will occur?" How can the general public have a useful or informed opinion on this highly technical and speculative question? The question brought answers of 57% "very likely," 34% "possible," and 7% "not likely." After giving these figures, the text obscures the careful tripartite posture of the question as stated by lumping the first two categories together in the following summary: "About 9 out of 10 consider it possible or very likely ..."

The other five questions asking for similar 25-year predictions (for example, a cure for the common forms of cancer) are not much better.

In sum, the public opinion surveys currently used in *Science Indicators* are poorly constructed and carelessly reported. By emphasizing remote and speculative uses of science, the thrust of the questions misrepresents the very nature of scientific method. (There are worse misrepresentations, for example, in a report for GAO (General Accounting Office), mistitled *Science Indicators: Improvements Needed in Design, Construction, and Interpretation*).

To summarize: Opinion surveys may attempt to reduce to numbers both nebulous opinions and other qualities not easily so reducible. It would be wiser if their use were restricted to those things which are properly numerical.

My own chief experience with other unhappy attempts to use mathematical ideas where they do not fit comes from studying many of the reports of the National Research Council (in brief, the NRC). I recently served for eight years as chairman of the Report Review Committee for this Council. This Council operates under the auspices of the National Academy of Sciences, which by its charter from the government is required to provide, on request, advice on questions of science or art. There are many such requests. Each year, to this end, the NRC publishes several hundred reports, aimed to apply scientific knowledge to various questions of public policy. Some of these policy questions are hard or even impossible of solution, so it may not be surprising that the desire to get a solution and to make it precise may lead to the use of quantitative methods which do not fit. This lack of fit can be better understood at the hand of some examples.

VI. Cost-Benefit and Regression

Before making a difficult decision, it may be helpful to list off the advantages and the disadvantages of each possible course of action, trying to weigh the one against the other. Since a purely qualitative weighing of plus against minus may not be objective (or at any rate can't be done on a computer), there has grown up a quantitative cost-benefit analysis, in which both the costs and the benefits of the action are reduced to a common unit—to dollars or to some other such "numeraire." The comparison of different actions and thus perhaps a decision between them can then be made in terms of a number, such as the ratio of cost to benefit.

In simple cases or for isolated actions this may work well; I am told that it did so function in some of its initial uses in decisions about plans for water resources. However, the types of decisions considered in NRC reports were usually not so straight-forward. I studied many such reports which did attempt to use cost-benefit analysis. In every such case which came to my attention in eight years, these attempts at quantitative cost-benefit analysis were failures.

In most cases, these failures could have been anticipated. Sometimes the intended cost-benefit analysis was not an actual numerical analysis but just a pious hope. For instance, one study tried to describe ways to keep clean air somewhere "way out west." In this case, there weren't enough dependable data to arrive at any numbers for either the costs or the benefits of that clean air. Hence the report initially included a long chapter describing how these costs and benefits *might* be calculated—although it really seemed more likely that there never would be data good enough to get dependable numbers for such a calculation.

There are also cost-benefit calculations which must factor in the value of the human lives which might be saved by making (or not making) this or that decision. In such cases, the value ascribed to one human life can vary by a factor of 10, ranging from one hundred thousand to one million dollars. Much of the variation depends on whether one gets the value of that life in terms of discounted future earnings or by something called implicit self-valuation of future satisfaction. However, I strongly suspect that whatever the method, there isn't any one number which can adequately represent the value of human life for such cost-benefit purposes. Our lives and our leisures are too various and their value (to us or to others) is not monetary. The consequence is that decisions which deal substantially with actions looking to the potential saving of lives cannot be based in any satisfactory way on cost-benefit analysis.

Another aspect of cost-benefit methods came to my attention just yesterday, in the course of a thesis defense. Cost-benefit methods attend only to gross measures, in a strictly utilitarian way, and give no real weight to the distribution of benefits (or of costs) between individuals.

Another striking example of the problems attending the use of cost-benefit analysis in policy studies is provided by a 1974 NRC study, "Air Quality and Automobile Emission Control," prepared for the Committee on Public Works, U.S. Senate. That committee was considering the imposition of various levels of emission controls on automobiles; it requested advice on the merits of such controls, and in particular wanted a comparison of the costs and the benefits of such control.

Some benefits of the control of automobile emissions are to be found in cleaner air and some in better health (less exposure to irritating smog). A number of studies of such health effects had been done; the NRC committee examined them all and considered all but one of them inadequate. The one adequate study was for students in a nursing school in the Los Angeles area. Each student carefully recorded daily discomforts and illnesses; these records were then correlated with the observed level of smog in Los Angeles. The results of this one study were then extrapolated by the NRC committee to the whole of the United States in order to estimate the health benefits of decreasing smog! It was never clear to me why Los Angeles is typical or how such a wide extrapolation can be dependable. Just as in the case of saving lives, the benefits of good health can hardly be reduced to numbers.

Some other difficulties with this particular study concern the use of regression, a mathematical topic with a considerable history. Mathematics deals repeatedly with the way in which one quantity y may depend upon one or more other quantities x. When such a y is an explicitly given function of x, the differential calculus has made extraordinarily effective use of the concept of a derivative $dy/dx = y'$; in the first instance, the use of the derivative amounts to approximating y by a linear function, such as $y = ax + c$, choosing a to be a value of the derivative y'. The number a then is units of y per unit of x and measures the number of units change in y due (at x) to a one-unit change in x. For certain purposes these linear approximations work very well, but in other cases, the calculus goes on to use higher stages of approximation — quadratic, cubic, and even an infinite series of successive powers of x.

But a variable quantity y involved in a policy question is likely to depend not just on one x, but on a whole string of other quantities x, z, and so on. Moreover, the fashion of this dependence can be quite complex. One approximation is to again try to express y as a constant a times x plus a constant b times z and so on—in brief to express y as a linear function

$$y = ax + bz + \cdots$$

with coefficients a, b, ... which are not yet known. Given enough data, the famous method of "least squares" will provide the "best" values of the constants a, b, ... to make the formula fit the given data. In particular, the coefficient a estimates the number of units change in y per unit change in x—holding the other quantities constant (if one can).

This process is called a multiple "regression" of y on x, z, ... This curious choice of a word has an explanation. It was first used by Galton in his studies of inheritance. He noted that tall fathers had sons not quite so tall—thus height had "regressed on the mean."

This technique of regression has been amply developed by statisticians and others; it is now popular in some cost-benefit analyses. For example, with the control of auto emission,

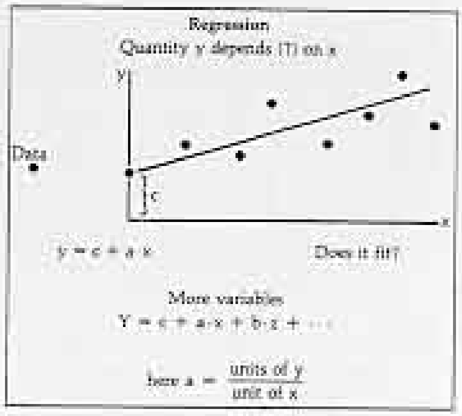


Figure 9

how does one determine the benefit of the resulting clean air?

Clean air cannot be purchased on the market, so the benefits of cleaner air might be measured by "shadow" prices found from property values, on the grounds that homes in a region where the air is clear should command higher prices than comparable homes where the air is thick. In the NRC study, the prices of houses in various subregions of greater Boston were noted and then expressed as a (linear) function of some thirteen different measured variables thought to influence these prices: Clean air, proximity to schools, good transportation, proximity to the Charles River, and so on. The constants in this linear expression of house prices were then determined by regression. In this equation, the coefficient a for the variable representing "clean air" (of units of dollars per measure of smog-free air) was then held to give the "shadow price" for clean air. The resulting shadow price from this and one other such regression was then extrapolated to the whole U.S.A. to give a measure of the benefit of cleaner air to be provided by the proposed auto emission control.

In Boston: House \$ = a (Smog) + b (Charles River) + ··· one dozen more

---

This is surely a brash attempt to get a number, cost what it may. In my considered judgment, the result is nonsense. It is not clear that buyers of houses monitor the clean air before they sign the mortgage. A nebulous (or even an airy) quantity said to depend on thirteen other variables is not likely to be well grasped by any linear function of those variables. Some variables may have quadratic effects, and there could be cross effects between different variables. That list of thirteen variables may have duplicates or may very well miss some variables which should be there. Moreover, the coefficients in that function are likely to be still more uncertain than the known costs the equation estimates. These coefficients are not shadow prices; they are shadowy numbers, not worthy of serious regard. They employ a mathematics which does not fit.

The difficulties which have been noted in interpreting the coefficients in some regressions are by no means new. For example, you can find them discussed with vigor and clarity in a text by Mosteller and Tukey, *Data Analysis and Regression*, kept here on permanent reserve in the Eckhart Library. I trust that such reserve has not kept it from the eyes of economists or other users of regression. What with canned formulas from other sources and fast computers, any big set of data can be analyzed by regression—but that doesn't guarantee that the results will fit!

I have not studied the extensive academic literature on cost-benefit analysis, but these and other flagrant examples of the misuse of these analyses in NCR reports leave me disquieted. Current political dogma may create pressure for more cost-benefit analysis. In Congress, the House is now considering a "Regulatory Reform Bill" which requires that independent and executive agencies of the government make a cost-benefit analysis before issuing any new regulation (except for those health and safety regulations required by law). It is high time that academicians and politicians give more serious thought to the limitations of such methods of analysis.

The future is inscrutable. However, people are curious, so fashion usually provides some method for its scrutiny. These methods may range from consultation with the Oracle at Delphi to opinion polls to the examination of the entrails of a sacrificial animal. Now, thanks to the existence of fast computers, some economists can scrutinize the future without entrailing such sacrifice. The short-term predictions by econometric models can be sold at high prices, though I am told that some of these models deliver more dependable short-term predictions when the original modeler is at hand to suitably massage the output figures.

At the NRC, my chief contact with projection was on a very much longer time scale—econometric projections of the energy future of the United States going forward for fifty years or more. This was done in connection with a massive NRC study called CONAES (for the Committee on Nuclear and Alternative Energy Systems). For this study, there was not just one econometric projection of energy needs, but a half dozen such models, with a variety of time horizons. Now projections for a span of forty or fifty years cannot possibly take account of unexpected events such as wars, oil cartels, depressions, or even the discovery of new oil fields. Since the present differs drastically from the past, there is little or no hope of checking a fifty-year projection against fifty years of actual past development. Consequently, this particular NRC study did not check theory against fact, but just theory against theory—by asking just how much agreement there was between the half-dozen models. It hardly seemed reasonable to me to conclude that agreement—even a perfect agreement — in the results of several fictive models can be of any predictive value. In the case of the CONAES report, there was even a proposal to use the thirty-five-year projection of those models to assess the future economic value of the breeder reactor. Such assessment breeds total futility. All told, despite the use of fast computers and multiple models, the ambiguities of the models being computed still leave the future dark and inscrutable.

Projections over time into an unknown future are not the only examples of policy-promoted projection of the unknown. Many other types of extrapolation can be stimulated—for example, extrapolation designed to estimate risks. Since it is claimed society has become more risk-averse, there is great demand to make studies of future risks, as in the reports of the NRC Committee on the Biological Effects of Ionizing Radiation (BEIR for short). The third report of this committee, a report commonly known as "BEIR III," dealt with extrapolation, another kind of projection. Data available from Hiroshima and Nagasaki give the numbers of cancers caused by high dosages of radiation. For present puropses one wants rather the effect of low doses, on which there are little or no data. To estimate this effect, one may assume that the effect E is proportional to dosage D—so that $E = kD$ for some constant k. Alternatively, on may assume that the effect is quadratic so that E depends both on D and on D^2. Then the curve giving E as a function of D is parabolic (Figure 10). The constants involved—such as the proportionality factor k—are then chosen to get the best fit of the line or the parabola to the high dosage data. The resulting formula is then used to calculate the effect at low dosage. Quite naturally, the linear formula and the quadratic one give substantially different results by this extrapolation; this is the cause of considerable controversy. Is the linear formula right? Does the choice of formula depend on the type of cancer considered? There is no secure and scientific answer to these pressing policy questions. In particular, the mathematical methods themselves cannot possibly produce an answer. Mathematical models such as these may be internally consistent, but that doesn't imply that they must fit the facts. Here, as in the case of regression, the assumption that the variables of interest are connected by a linear equation is gratuitous and misleading.

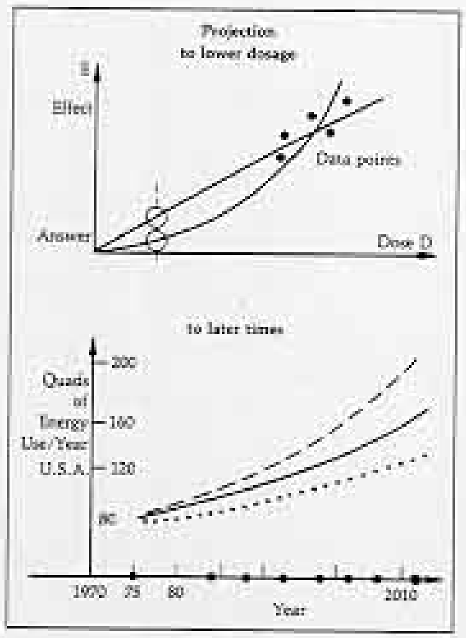


Figure 10

That BEIR III report deals with just one of many different kinds of risks that plague mankind. There are many others that might be estimated, by extrapolation or otherwise. From all these cases there has arisen some hope that there might be effective general principles underlying such cases–and so constituting a general subject of "risk analysis." The hope to get at such generality may resemble the process of generalization so successful in mathematics, where properties of numbers have been widely extended to form the subject of number theory and properties of specific groups have led to general group theory. However, I am doubtful that there can yet be a generalized such "risk analysis" — and this I judge from another current NRC report.

This report arose as follows: The various public concerns about risks were reflected in Congress, so a committee of the Congress instructed the National Science Foundation (NSF) to establish a program supporting research on risk analysis. The NSF, in its turn, did not know how to go about choosing projects in such a speculative field—so it asked the National Research Council for advice on how to do this. The NRC, again in turn, set up a committee of experts on risk analysis. This committee, in its turn, prepared a descriptive report on risk analysis "in general." The report also commented on specific cases of risk analysis. For example, there were extensive comments on the BEIR III report—but these comments did not illuminate the BEIR III problem of extrapolation and made no other specific suggestions. The report had little of positive value to help the NSF decide which projects in risk analysis to fund. Such a general study of risk analysis is clearly interdisciplinary, but I must conclude that it is not yet disciplined.

These and other examples of unsatisfactory reports may serve to illustrate the confusion resulting from questionable uses of quantitative methods or of mathematical models. But why are there so many cases of such confusion? Perhaps the troubled history of that report on risk analysis is typical. A practical problem appears; many people are concerned, and so is the Congress or the Administration. Since the problem is intractable, but does involve some science, it is passed on to the scientists, perhaps to those at the NRC. Some of these problems can be—and are—adequately treated. For others there is not yet any adequate technique—and so those techniques which happen to be available (opinion surveys, cost-benefit analysis, regression, projection, extrapolation, decision analysis, and others) get applied to contexts where they do not fit. Confusion arises when the wrong idea is used, whether for political reasons or otherwise.

There are also political reasons for such confusion. Our representatives, meeting in that exclusively political city of Washington, represent a variety of sharply different interests and constituencies. To get something done, a compromise must be struck. This happens in many ways. One which I have seen, to my sorrow, is the adjustment of the onerous and bureaucratic regulations of the OMB (Office of Management and Budget) about cost principles for universities. Their Circular A-21 now requires faculty members to report the percentage distribution of their various university activities, with results to add up to 100%, on a "Personnel Activity Report Form" (PAR!). Such numbers are meaningless; they are fictions fostered by accountants. Use of such numbers makes for extra paperwork—but it also tends to relocate some control of scientific research from universities to the government bureaucrats. For A-21, there was recently a vast attempt at improvement, combining all parties: the government bureaucrats, their accountants, university financial officers, and a few faculty. What resuted? A compromise, and not a very brilliant one.

Thus government policy, when it requires scientific advice on matters that are intrinsically uncertain, is likely to fall into the government mold: compromise. And that, I believe, is a source of confusion.

VIII. Fuzzy Sets and Fuzzy Thoughts

The misuse of numbers and equations to project the future or to extrapolate risks is by no means limited to the National Research Council. Within the academic community itself there can be similar fads and fancies. Recently I have been reminded of one curious such case: The doctrine of "fuzzy" sets.

How can a set be fuzzy? Recall that a set S is completely determined by knowing what things x belong to S (thus $x \in S$) and what things do not so belong. But sometimes, it is said, one may not know whether or not $x \in S$. So for a fuzzy set F one knows only the likelihood (call it $\lambda(x)$) that the thing x is in the fuzzy set F. This measure of likelihood may range from 0 (x is certainly not in F) all the way to 1 (x is certainly in F). Now I might have said that $\lambda(x)$ is the probability that x is in F, to make this definition a part of the well-established mathematical theory of probability. The proponents do not so formulate it, because their intention is different and much more ambitious: Replace sets everywhere by fuzzy sets!

By the grand set-theoretic doctrine, every mathematical concept can be defined in terms of sets, hence this replacement is very extensive. It even turns out that many mathematical concepts can be fuzzed up in several ways, say, by varying the fuzzy meaning to be attached to the standard set-theoretic operations (intersection, union, etc.) of the usual Boolean algebra of sets. And so this replacement doctrine has already produced a considerable literature: on fuzzy logic, fuzzy graphs, fuzzy pattern recognition, fuzzy systems theory, and the like. Much of this work carries large claims for applications of this fuzzy theory. In those cases which I have studied, none of the applications seem to be real; they do not answer any standing problems or provide any new techniques for specific practical situations. For example, one recent book is entitled *Applications of Fuzzy Sets to Systems Analysis*. The actual content of the book is a sequence of formal fuzzy restatements of standard mathematical formulations of materials on programming, automata, algorithms, and (even!) categories, but there is no example of specific use of such fuzzy restatement. One reviewer (in *Mathematical Reviews*) noted a "minimal use or lack of instructive examples—the title of the book purports applications." Another more recent book on fuzzy decision theory states as one of its six conclusions, "It is a great pity that there exist only very few practical applications of fuzzy decision theories, and even practical examples to illustrate the theories are scarce." This leads me to suspect that the initially ingenious idea of a fuzzy set has been overdeveloped in a confusing outpouring of words coupled with spurious claims to importance.

There are other examples—cybernetics, catastrophe theory— where an originally ingenious new idea has been expanded uncritically to lead to meaningless confusion.

IX. Compromise Is Confusing

But enough of such troubling examples of confusion. Let me summarize where we have come. As with any branch of learning, the real substance of mathematics resides in the ideas. The ideas of mathematics are those which can be formalized and which have been developed to fit issues arising in science or in human activity. Truth in mathematics is approached by way of proof in formalized systems. However, because of the paradoxical kinds of self-reference exhibited by the barn door and Kurt Gödel, there can be no single formal system which subsumes all mathematical proof. To boot, the older dogmas that "everything is logic" or "everything is a set" now have competition–"everything is a function." However, such questions of foundation are but a very small part of mathematical activity, which continues to try to combine the right ideas to attack substantive problems. Of these I have touched on only a few examples: Finding all simple groups, putting groups together by extension, and characterizing spheres by their connectivity. In such cases, subtle ideas, fitted by hand to the problem, can lead to triumph.

Numerical and mathematical methods can be used for practical problems. However, because of political pressures, the desire for compromise, or the simple desire for more publication, formal ideas may be applied in practical cases where the ideas simply do not fit. Then confusion arises — whether from misleading formulation of questions in opinion surveys, from nebulous calculations of airy benefits, by regression, by extrapolation, or otherwise. As the case of fuzzy sets indicates, such confusion is not fundamentally a trouble caused by the organizations issuing reports, but is occasioned by academicians making careless use of good ideas where they do not fit.

As Francis Bacon once said, "Truth ariseth more readily from error than from confusion." There remains to us, then, the pursuit of truth, by way of proof, the concatenation of those ideas which fit, and the beauty which results when they do fit.

If only Longfellow were here to do justice to the situation:

Tell Me Not in Fuzzy Numbers

In the time of Ronald Reagan
Calculations reigned supreme
With a quantitative measure
Of each qualitative dream

With opinion polls, regressions
No nuances can be lost
As we calculate those numbers
For each benefit and cost

Though his budget will not balance
You must keep percents of time
If they won't sum to one hundred
He will disallow each dime.

References

"Air Quality and Automobile Emission Control." A report prepared by the Commission on Natural Resources, National Academy of Sciences for the Committee on Public Works, U. S. Senate. September 1974. Volume 4. *The Costs and Benefits of Automobile Emission Control.*

Eilenberg, S. and Mac Lane, S. "Group Extensions and Homology," *Annals of Math.* 43 (1941), 758-831.

Eilenberg, S. and Mac Lane, S. "General Theory of Natural Equivalences" (category theory), *Trans. Am. Math. Soc.* vol. 28 (1945), 231-294.

Energy Modeling for an Uncertain Future: Study of Nuclear and Alternative Energy Systems. A series: Supporting Paper #1, 225pp. National Academy Press, 1978.

Feit, Walter and Thompson, John G. "Solvability of Groups of Odd Order," *Pacific J. Math.* 13 (1968), 775-1029.

Kickert, Walter J. M. Fuzzy Theories in Decision-making: A Critical Review. Martinus Nijhoff, Social Sciences Division, Leyden, 1978. 182 pp. Reviewed in *Mathematical Reviews* vol. 81f (June 1981), #90006.

Ladd, Everett C. and Lipset, Seymour M. The 1977 Survey of the American Professoriate.

Lang, Serge. The File: A Case Study in Correction. 712 pp. Springer Verlag, New York, 1981.

Lawvere, R. W. "Toposes, Algebraic Geometry, and Logic," *Springer Lecture Notes in Math.* No. 274 (1972).

Mosteller, F. and Tukey, John W. Data Analysis and Regression. 586 pp. Addison-Wesley, Reading, Mass. (1977).

Negoita, C. V. and Ralescu, D. A. Applications of Fuzzy Sets to Systems Analysis, 1975. John Wiley and Sons, New York. 191pp. Reviewed in *Mathematical Reviews* vol. 58 (1979), #9442a.

Science Indicators: Improvements Needed in Design, Construction, and Interpretation. Report by the Comptroller General of the United States. PAD 79-35, September 25, 1979.

Whitehead, A. N. and Russell, Bertrand. *Principia Mathematica*, vol.1, 2nd edition. Cambridge University Press, 1925. 674 pp.

The Ryerson Lecture was given April 20, 1982 in the Glen A. Lloyd Auditorium of the Laird Bell Law Quadrangle.

The Nora and Edward Ryerson Lectures were established by the trustees of the University in December 1972. They are intended to give a member of the faculty the opportunity each year to lecture to an audience from the entire University on a significant aspect of his or her research and study. The president of the University appoints the lecturer on the recommendation of a faculty committee which solicits individual nominations from each member of the faculty during the winter quarter preceding the academic year for which the appointment is made.

The Ryerson Lecturers have been:

1973-74: John Hope Franklin, "The Historian and Public Policy"
1974-75: S. Chandrasekhar, "Shakespeare, Newton, and Beethoven: Patterns of Creativity"
1975-76: Philip B. Kurland, "The Private I: Some Reflections on Privacy and the Constitution"
1976-77: Robert E. Streeter, "WASPs and Other Endangered Species"
1977-78: Dr. Albert Dorfman, "Answers Without Questions and Questions Without Answers"
1978-79: Stephen Toulmin, "The Inwardness of Mental Life"
1979-80: Erica Reiner, "Thirty Pieces of Silver"
1980-81: James M. Gustafson, "Say Something Theological!"

ARTICLE

DIGITAL MAPS AND DATA BASES: AESTHETICS VERSUS ACCURACY *

Robert F. Austin, Ph.D.
President
Austin Communications Education Services
28 Booth Boulevard
Safety Harbor, FL 34695-5242
Past-president, AM-FM International

I. Introduction

Of the many courses lectured by Immanuel Kant at the University of Königsberg, legend has it that one of the most frequently offered was a course on natural philosophy (that is, physical geography). It was argued that individuals could acquire understanding through three distinct perspectives: the perspective of formal logic and mathematics, the perspective of time (history), and the perspective of space (geography). The last of these — the perspective of space — acknowledges the importance of distance, site characteristics, and relative location in describing the relationships among several objects or facilities.

Maps are the primary means of representing such relationships. Maps are analytical tools which depict spatial relationships and portray objects from the perspective of space. The power of maps rests on their synoptic representation of complex phenomena. To paraphrase the Confucian wisdom, a map is worth a thousand words.

> "Recently it has become common to convert spatial phenomena to digital form and store the data on tapes or discs. These data can then be manipulated by a computer to supply answers to questions that formerly required a drawn map ... This stored geographic information is referred to as a [data base]." [1]

The maps produced from such data bases are termed digital maps. Computerized data bases, which may be queried and used by several people simultaneously, and digital maps are of immense value to engineers, comptrollers, planners, and managers. The combination of a digital map and data base is worth a thousand "mega-words."

The advantages of digital maps over manually drafted maps are most apparent in situations of frequent growth or change. Among these advantages are the ease and speed of revision and the fact that special purpose maps can be produced in small volume at reasonable cost. Moreover, digital maps offer greater precision in representation and analysis. "As more governmental bodies [and other agencies] expend the necessary one-time capital investment, and begin to reap the vast rewards of computer-assisted record and map keeping, others are likely to follow quickly."[2]

II. Basic Issues

After an agency or firm has decided to convert its manual records to digital map and data base form, several issues must be addressed.

The first, and most important question the agency must answer is related to the source documents to be used by the mapping firm. In general terms the choice is between cartographic sources and mechanically drafted sources.

Cartography is defined as:

> "The art, science and technology of making maps, together with their study as scientific documents and works of art. In this context maps may be regarded as including all types of maps, plans, charts, and sections, three-dimensional models and globes representing the Earth or any celestial body at any scale."[3]

In particular, cartography is concerned with the accurate and consistent depiction on a flat surface of activities occurring on a sphere.

It is not possible to duplicate, without distortion, the features on the surface of a sphere on any object other than a sphere. A surface of constant positive curvature may be represented on a surface of zero curvature only if distortion is introduced in the representation. As a simple illustration of this fact, consider the problem of "flattening" an orange peel: it will tear. If the orange was made of rubber, it would be possible to flatten it without tearing, but not without distortion of another kind — a topological transformation.

The methods by which cartographers represent the surface of the earth on a flat piece of paper are known as map projections. For any particular purpose, the selection of a particular projection (transformation) is based on the properties of a sphere that the projection loses or retains. Every method of mapping large areas is affected, whether it is continuous mapping or facet mapping. No coherent, distortion-free transformation exists, nor, given the theorems of mathematics, can it ever exist. [4, 5] However, cartographers can identify projections that suit a client's particular purpose.

Quite often appropriate cartographic source documents already are available to a public utility and mapping firm team. Indeed, such sources may have served as the base for the construction of existing records. In other cases, it may be necessary for the mapping firm to perform an aerial photographic survey and to translate these photographs into cartographic documents — a process known as photogrammetry.

It is also possible to produce maps from non-cartographic sources such as tax assessor sheets. Certainly the most common non-cartographic sources are mechanically drafted cadastral maps and engineering drawings or plans. These documents have been defined by the International Association of Assessing Officers:

> map, cadastral — A map showing the boundaries of subdivisions of land, usually with the bearings and lengths thereof and the areas of individual tracts, for purposes of describing and recording ownership.
>
> map, engineering — A map showing information that is essential for planning an engineering project or development and for estimating its cost. An engineering map is usually a large-scale map of a comparatively small area or of a route. [6]

Although such drawings have some value for small area design, engineering, and planning purposes, there are a number of problems associated with their use as source documents for large area mapping. The most critical of these is related to accuracy; tax assessor sheets in the United States, for example, are designed to be used as indices only and are subordinate to actual legal descriptions. They are highly stylized and, despite their appearance and name, highly inaccurate in terms of geographic placement.

Small plans "look" correct primarily because they correspond to the limited range of vision of human beings at ground level. However, these documents also suffer from the transformation problem. Non-uniform, interpretive, subjective corrections by a draftsman make this problem appear to vanish on individual sheets. But such corrections preclude accurately merging sheets for a large area.

In the language of the philosophy of science, the distinction is one between an iconic model and a symbolic model. An iconic model (the mechanically drafted plan) is designed to look, in some metaphorical fashion, like the object of study. Often, the closer the similarity in appearance, the less valuable the model for analytical purposes. A symbolic model (the map) is designed to facilitate quantitative measurements of characteristics of interest to analysts, managers, and engineers.

A second issue that must be considered by public utilities is the use to which the digital maps will be put. This will determine the type of output the mapping firm will generate. This also will determine the accuracy levels needed. [7] In general terms, the types of output products correspond to the types of input products: maps and plans. [8] In our experience, public utility clients generally have expressed a preference for digital maps related to spatial relationship data bases because they facilitate the more accurate analysis of physical plant attributes and distributions over a large area in a geographic information system.

The primary purpose of most digital map and data base conversion work is to provide a means to manage corporate assets. Often the actual maps produced are used for index only, not for scaled or direct measurement. This is in part a function of the distortion inherent to any mechanical production or reproduction process, in part a function of the demonstrated superiority of a fully digital, displayable linked-attribute data base management system (see Section 5), and in part a function of the distortion inherent to all map projections (the transformation problem previously discussed.)

In some cases, an agency may wish to construct a geographic data base that will support a computerized plan generation and facilities management system. As an example, consider the case where facility data will be superimposed on a merged cadastral and land base. The data base must guarantee the geographic locations of features and their connectivity, relationships, and other characteristics. The final digital plan and data base may include information on street, road, and highway names, centerlines, and rights-of-ways; political, legal, and natural boundaries; township, range, and section lines; river, stream, and creek centerlines and names; and legal lot and parcel lines and numbers, among other data. [9]

III. Map Production

Regardless of the type of source document or output product, several stages in cartographic production remain relatively constant. These are considered first in a general manner and then as they apply to digital map production per se.

First, we must define the purpose and accuracy standards of the map. For example, will the map be used for scaled measurement? Or will the map be used as an index? Second, we must identify the features and activities to be mapped. The nature of these features will influence the amount of detail appropriate for the base map and the finished map. The strength of a map may be diminished by displaying too much detail.

Third, we must prepare or obtain a land base or base map. In this regard, it is important to consider the variety of map projections and coordinate systems available for particular tasks. Using a widely accepted system such as the UTM grid or latitude and longitude coordinates has a number of advantages, including ease of data exchange and reduced production time and cost.

The next step is to collect and compile the data to be mapped. The basic rule is to compile data at the most detailed level of measurement possible and to aggregate the data only at later analytical stages. Finally, we must design and construct the map. This is a two step process that involves: (a) the design of symbols, patterns, legends, and other cartographic devices, and (b) the location and actual placement of the features and activities.

IV. Digital Maps

As in traditional cartography the first step in constructing a digital map is to establish accuracy levels and to determine which attributes should be displayed and which should simply be stored.

A displayable linked-attribute data base system (discussed below) allows for the construction of a fully digital geographic information system for data management, as well as for the construction of index and general route maps. For such maps, placing items of plant "on the right side of the street" generally is adequate: in the real world, a utility pole 60 feet tall is clearly visible at an intersection. The critical attributes of each item of plant appear as numbers or words on the map and also as manipulable information in the data base. On the other hand, if the map is to be used for scaled measurement, accurate placement of items of plant is paramount. It should be noted that this second approach implies considerable supplementary manual adjustment and therefore substantially higher production costs.

The next question is the method of land base construction. Land bases, or base maps, may be constructed in a variety of ways. It may be possible to develop an accurate land base by digitizing existing source documents. The information may be captured by board digitizing vectors (line strings and endpoints) or by raster scanning. If the quality of the source documents is high, these methods are extremely cost effective.

If the quality of the available source documents is unknown or suspect, it is common to conduct an aerial photographic survey and to compile the photographs into a "model" of the region (the air photos are rectified to generate a plane view of the photographed region similar to a map projection). These models are then stereo digitized and used as highly accurate source documents for land base construction using the digitizing or scanning methods noted above. (± 10 feet accuracy is standard, but ± 1 foot accuracy is possible.) Although more expensive than working from existing source documents, such an approach guarantees extreme accuracy. Moreover, the end product often has resale value which will offset the initial cost incurred by the end user.

Data sources also may be combined to form a hybrid land base. For example, individual assessor sheets at a wide variety of scales can be overlaid on a stereo digitized base. The stereo digitized street centerline network can be modified to agree with cadastral maps so the assessor data will scale properly and satisfy aesthetic criteria.

Because such a hybrid is, by definition, unique, it is appropriate to discuss in detail at the outset of the project the problems likely to be encountered in production. Disadvantages of such an approach include the substantial cost to make the map "look" like the source documents, many difficult production problems (e.g., warp of cadastral data and fitting cadastral information to an accurate land base), and the volume of source documents required.

Nevertheless, some utilities wish to use a hybrid approach because it gives them a product that is internally acceptable from an aesthetic viewpoint. We have encountered many situations where end users are uncomfortable with the computer-plotted version of a geographic data base because it does not "look" like the product they have used for many years. The issue of user acceptance is of critical importance to the ultimate success of a conversion project. The cost and problems associated with a hybrid approach may be justified in the long run because the end user is comfortable with the "look" of the product. However, care must be taken to avoid simply computerizing existing problems.

The process of generating a hybrid combines land base construction and data compilation. When a data base management system approach is used, the distinction between these two production processes is much clearer. After a land base is assembled, a decision is made to define some selection of "attributes" of facilities or items of plant as displayable. Displayable attributes are those attributes that actually will be plotted on a map. Other attributes may be stored in the data base, but not, as a matter of course, be displayed on maps. (Information of interest to comptrollers may be of little use to engineers. Conversely, attributes that facilitate engineering procedures may be unimportant to comptrollers.)

Once agreement on the attributes to be displayed is reached, the data are coded and laid out on the source document. Some of the information will be recorded interactively at a board digitizing station. Other information will be keypunched and bulk loaded at the construction phase. After construction, updating normally is performed interactively. One significant advantage of a data base management system approach is its ability to grow or change with technical advances.

> "Maps today are strongly functional in that they are designed, like a bridge or a house, for a purpose. Their primary purpose is to convey information or to 'get across' a geographical concept or relationship ... The mapmaker is essentially a faithful recorder of given facts, and geographical integrity cannot be compromised to any great degree. Nevertheless, the range of creativity through scale, generalization, and graphic manipulation available to the cartographer is comparatively great."[10]

Given its pragmatic character, it may be surprising to learn that the physical appearance of a map is a common point of disagreement. Most frequently such disagreement arises because different sets of aesthetic principles have been applied by the client and the mapping firm.

The general question of aesthetics is not at all simple; as the art historian Ivins has argued, the notion of simple geometric relationships is not invariant in aesthetic assessments. [11] Indeed, aesthetic issues are involved in both the creation and the appreciation or perception of a work of art. Within cartography, the term "aesthetics" is reserved for consideration of the placement of elements such as compass rose, legends, and scales; the balance of these elements vis-a-vis the map object; the selection of type faces from the range of standard or customary fonts; and similar elements of visual display. To equate "correct" with cartographic aesthetics is inappropriate, except in the limited sense that some features are required by cartographic convention (e.g., italic type fonts for bodies of water). Styles as to what is "correct" from a creative standpoint also vary from one academic discipline to another: the geographer prefers a fine-line drawing while the urban planner uses heavy lines to focus attention and the landscape architect employs pictorial symbols. Differences in styles of map creation help to condition the manner in which maps are appreciated by consumers; that is, aesthetics is more than simply giving the consumer what he is used to. For example, many public utilities have developed a sense of aesthetics conditioned by experience with manually drafted, subjective, and highly symbolic plans. Unless discussed at the beginning of a conversion project this point can become most difficult, because a mapping firm must assume that map accuracy takes precedence over map symbology, visual appearance, and the superficial aesthetics of the perception of appearance. Often in cases of this sort, the aesthetics of appreciation of the consumer give way to concerns of accuracy on the part of the mapping firm which in turn may rest on the aesthetics of cartographic creation.

V. Computerized Data Bases

> "Computer systems are increasingly used to aid in the management of information, and as a result, new kinds of data-oriented software and hardware are needed to enhance the ability of the computer to carry out this task. [Data base systems are] computer systems devoted to the management of relatively persistent data. The computer software employed in a data base system is called a data base management system (DBMS)." [12]

Of the several methods of classifying data bases used by software engineers, one most important dichotomy is that between network (hierarchical) and relational data bases. [13] Certainly the most useful approach for many users is the relational data base, because this approach permits a larger variety of queries. Regardless of the approach, the method of manipulating the data base remains a critical issue.

As noted in the quotation of Blasgen [12], the software used with a data base is known generically as a data base management system (DBMS). Such a system generally will have provisions for data structure definition as well as for data base creation, maintenance, query, and verification. Blasgen observed that in 1981 an estimated 50 companies were marketing 54 different DBMS packages. [14]

In our experience, as noted earlier, most public utility clients prefer working with a "displayable linked-attribute" DBMS. This term describes a system in which selected attributes or characteristics of the company's physical plant are stored in the data base, where they may be manipulated by users and also displayed on the digital map. For example, the age, length, size, and identification number for a piece of cable may be stored and displayed. Because the length of cable in a given span is known from installation and stored in the data base, scaled measurement is unnecessary.

A second approach to building and maintaining a data base is termed the "hybrid" approach. Consider the following example.

Analytics technicians select National Geodetic Survey monuments and photo identifiable points which provide a network for accurate control of the area. On-site field survey crews accurately survey these points and target them for aerial photography. After the flight, analytics technicians assemble these points into an accurate control network and place them in a digital file. Using the network file, features defined in contract specifications are stereo digitized from the photography in a digital format. The major features are portrayed in the form of a detailed centerline network of interstate highways, public roads and private roads. The file then is divided into facets (corresponding to individual maps) and plotted at a scale of 1:100'.

Tax assessor sheets and the 1:100' stereo digitized centerline plots are joined at the next production phase. The 1:100' plot is overlaid on individual assessor sheets, intersections are held for control, and the stereo digitized road centerline network is adjusted to match the cadastral maps. The revised centerlines and additional features such as rights-of-ways, lot lines, boundaries, and text are board digitized.

The product is the best of the digital and mechanical worlds. It is accomplished in digital format so future modifications can be made easily, and it is aesthetically pleasing — it looks like an engineering drawing or cadastral map. The information is accurate, the data can be scaled, and bearings can be extracted.

The success of endeavors of this type depends on an excellent vendor and client relationship. For such a project, it is recommended that a test or pilot study in a pre-selected area be completed prior to final contract negotiations. It is the responsibility of the client to define his needs as accurately as possible and convey these requirements in understandable language to the vendor. Vague terminology and ambiguous specifications can compound production problems. The vendor, based on his background and experience, must make meaningful suggestions to the client as early in production as possible as alternative options may be considered.

VI. User Community

In summary, a comprehensive digital map and data system has many advantages. In order to fully realize these advantages, users must consider and resolve several questions related to actual needs and aesthetic conditioning. Users should understand that digital maps and data bases constructed using highly accurate aerial photographic source documents will not, as a rule, look like familiar graphic products. They must consider the distinctions between mechanically drafted products and cartographic products and decide at the outset of the project whether they are comfortable with the graphic specifications.

Close communication between the utility company and the mapping firm is critical. The more carefully specified the project is at the beginning, the fewer the changes that will be required. Changes in specifications made during production, no matter how trivial they appear, generally affect production schedules and costs adversely.

Thus, the creation of a digital map involves not only the mastery of current technology, in order to produce an "accurate" map, but it also involves an awareness of aesthetics, as well. Attention to aesthetics, as appreciation of the map by the consumer will ensure a satisfied client; to this end, considerable education of the client with attention to close communication is appropriate. At the other end, the mapping firm needs to consider the aesthetics of map creation. When the aesthetics of creation help to guide the choice of technology, an accurate and satisfying digital map is generally the end product.

References

1. Robinson, A., R. Sale and J. Morrison (1978), *Elements of Cartography* (4th edition), New York: John Wiley, p. 4.
2. Robinson, A., et al., p. 272.
3. International Cartographic Association (Commission II, E. Meynen, Chairman) (1973), *Multilingual Dictionary of Technical Terms in Cartography*, Wiesbaden: Franz Steiner Verlag GMBH.
4. P. W. McDonnell, Jr. (1979), *Introduction to Map Projections*, New York and Basel: Marcel Dekker, Inc.
5. H. S. M. Coxeter, (1974), *Projective Geometry*, (second edition), Toronto: University of Toronto Press.
6. American Congress on Surveying and Mapping and American Society of Civil Engineers, Joint Committee (1978), *Definitions of Surveying and Associates Terms* (rev.), p. 101.
7. "Symposium on the National Map Accuracy Standard," *Surveying and Mapping*, 1960, v.20, n.4, pp. 427-457, and M. M. Thompson and G. H. Rosenfield, "On Map Accuracy Specifications," *Surveying and Mapping*, 1971, v.31, n.1, pp. 57-64.
8. Cuff, D.J. and M.T. Mattson (1982), *Thematic Maps: Their Design and Production*, New York and London: Methuen.
9. For additional discussion of this approach, see Easton, C.H. (1975), "The Land Records Information System in Forsyth County, North Carolina," pp. 261-267 in *International Property Assessment Administration*, Chicago: International Association of Assessing Officers.
10. Robinson, A., et al., p.7. See also: Amheim, R. (1976), "The Perception of Maps," *The American Cartographer*, 1976, v.3, n.1, pp. 5-10.
11. Irvins, W. Jr. (1964), *Art and Geometry: A Study in Space Intuitions*, New York: Dover (reprint of 1949 Harvard University Press edition).
12. Blasgen, M. W., "Data Base Systems," *Science*, 1982, v.215, 12 February.
13. Codd, E., "Relational Data Base: A Practical Foundation for Productivity," *Communications* of the ACM, 1982, v.25, n.2.
14. Blasgen [12].

* Acknowledgments

An earlier version of this essay appeared as a Chicago Aerial Survey Production Report. Comments on the previous version by Mr. Harold Flynn are gratefully acknowledged, as is the support provided by Geonex Corporation during its preparation. The author wishes to thank an anonymous referee for comments useful in preparing the current version.

FEATURES

i. Press Clipping

Science, November 29, 1991, Vol. 254, No. 5036, copyright, the American Association for the Advancement of Science. Many thanks to Joseph Palca at *Science* for his continuing interest in online journals. The citation appeared in "Briefings" and is entitled "**Online Journals**," by Joseph Palca.

Online Journals

"When the AAAS and OCLC Online Computer Library Center announced the scheduled debut next year of their new journal—*The Online Journal of Current Clinical Trials* — they said it would be the world's first peer-reviewed, online science journal (*Science*, 27 September, p. 1480). Since then, two other such journals have made their presence known to *Science*. They are *Solstice: An Electronic Journal of Geography and Mathematics*, published by Sandra Lach Arlinghaus of the Institute of Mathematical Geography in Ann Arbor, Michigan, and *Flora Online* , published by Richard H. Zander, curator of botany at the Buffalo Museum of Science. Both have been around for about 2 years and are available free over several popular research computer networks."

NOTE: Readers wishing to contact Richard Zander can do so at bitnet address:
VISBMS@UBVMS

ii. Word Search Puzzle

The point of this puzzle is to develop familiarity (dispelling fear) with a selection of words possibly not familiar to student readers. The words are embedded in the jumble of letters below; not all letters in this array are part of a word in the list and other words may appear in the puzzle. Words from the list may be written from left to right, from right to left, from top to bottom, from bottom to top, or diagonally (in any direction). Solution is on the last page of this issue.

WORDS IN THE PUZZLE

Algorithm
Asymptote
Azimuthal
Circumpolar
Circumscribe
Conic
Converge
Curvature
Cylindrical
Divergent
Equatorial
Equidistant
Equinox
Exponent
Fractal
Gnomonic
Graticule
Integral
Inverse
Jacobian
Lambert
Latitude
Logarithm
Longitude
Matrix
Mercator
Meridian
Norm
Oblique
Orthogonal
Parallel
Polyconic
Projection
Rotation
Solstice
Stereographic
Tangent
Translation
Vector

P	O	R	T	H	P	O	L	Y	C	O	N	I	C	I	H	P	A	R	G	O	E	R	E	T	S
E	L	O	N	G	I	T	U	D	E	E	Q	U	I	D	I	S	T	A	N	T	N	E	B	C	C
Q	A	L	G	O	R	I	T	H	M	L	O	D	I	V	E	R	G	E	N	T	G	G	I	U	V
U	H	S	L	G	H	T	M	I	H	P	R	C	Y	M	S	A	N	E	G	R	R	P	R	E	R
A	T	E	Y	R	M	E	T	O	T	P	M	Y	S	A	P	L	N	U	O	A	R	V	C	E	O
T	U	L	A	M	B	E	R	T	I	A	C	I	N	O	M	O	N	G	T	T	A	T	S	A	T
O	M	R	O	N	P	T	R	B	R	E	N	N	R	O	P	P	X	I	C	T	O	I	M	L	A
A	I	X	M	R	T	T	P	M	A	S	C	G	C	X	C	M	C	A	U	R	R	T	U	N	T
L	Z	R	S	A	T	Z	O	L	G	C	M	D	E	N	U	U	C	R	A	U	N	E	C	O	I
R	A	T	F	E	T	R	S	T	O	M	U	Q	Q	N	L	C	E	R	F	E	G	L	R	A	O
I	R	C	V	R	E	R	G	N	L	E	U	V	U	E	T	R	A	N	S	L	A	T	I	O	N
P	O	O	I	T	A	R	I	S	T	I	L	C	A	T	O	I	N	M	U	T	H	A	C	Z	I
M	T	N	O	R	E	C	X	X	N	W	F	A	T	B	N	C	A	J	I	O	C	A	R	N	N
E	A	V	E	R	D	A	T	O	F	J	A	C	O	B	I	A	N	T	G	E	R	G	T	A	V
T	C	E	T	S	A	N	X	A	R	S	X	A	R	R	T	G	U	O	A	N	R	E	V	B	E
G	R	R	A	M	R	S	I	A	L	X	S	E	I	P	R	D	N	D	G	E	G	A	R	V	R
R	E	G	E	C	I	T	S	L	O	S	T	E	A	Z	E	A	M	U	T	R	H	A	L	S	S
A	M	E	R	I	D	I	A	N	Y	L	A	A	L	E	L	L	A	R	A	P	R	A	L	T	E
I	S	G	A	L	R	N	O	I	T	C	E	J	O	R	P	O	B	L	I	Q	U	E	C	A	L

iii. Software Briefs — Brief descriptions of software provided by the creator. Look for reviews of the software in subsequent issues of *Solstice* . The Institute of Mathematical Geography (IMaGe) makes no claim as to the accuracy of the statements made by the creators; the appearance of their comments in *Solstice* is **not** an endorsement, either direct or indirect, of the product by IMaGe or by anyone associated with either IMaGe or *Solstice*. These "Briefs" are simply presented as a way for software creators to share infomation, in an e-journal, with other possibly interested parties.

a. RangeMapper TM — version 1.4. Created by Kenelm W. Philip, Tundra Vole Software, Fairbanks, Alaska. Program and Manual by Kenelm W. Philip; commentary below from the Manual. "A utility for biological species range mapping, and similar mapping tasks in other fields." FNKWP@ALASKA

"RangeMapper TM is a Macintosh mapping utility program designed specifically for the field or museum biologist who wants to be able to produce, rapidly and easily, species range maps for various organisms. The program may also be used for mapping other kinds of data, in medical, sociological, geological, geophysical, biological, etc. applications.

The program is aimed at people whose mapping needs cover sizable areas. The most accurate data files in the map base are derived from the CIA mapping files, which are suitable for displaying regions down to 20-30 miles or so in linear extent without showing a 'polygon' effect from data point spacing. The data files also include the Micro World Data Bank files, which are suitable for mapping on a whole-world scale and down to regions perhaps 500 miles in linear extent.

On a Macintosh II or other machines in the Macintosh II family, most maps will be plotted in well under one minute. Once plotted, they can be saved to disk and re-loaded as needed. In either case, data from properly formatted latitude/longitude textfiles can be read by the program and plotted to the maps. If your species data (including lat/long coordinates for collecting sites) are stored in a computer database, it should be easy to arrange to export, for any given species, a textfile of lat/long coordinates that RangeMapper can read directly — thus obtaining your range maps in one step from your database.

In conjunction with the word processor NISUS, RangeMapper may be used as a map-based visual interface to a text database.

The current version (1.4) of RangeMapper is set up for world mapping in six projections — north polar azimuthal, cylindrical, Mercator, orthographic, stereographic, and Lambert azimuthal equal-area. The north polar azimuthal maps are quite usable down to the southern limits of the lower 48 states and equivalent latitudes in Eurasia, and are excellent for higher-resolution mapping of Alaska. The cylindrical and Mercator projections can show the entire world (barring the extreme polar regions in Mercator), centered at any longitude. The last three projections show up to one hemisphere, which may be centered at any point on the earth's surface.

The map data files from the Micro World Data Bank cover the entire world. The only CIA file presently incorporated into RangeMapper is the Alaska file (approximately 150,000 points for coastlines, islands, rivers, and lakes). The entire continent of North America will be added from CIA files later, permitting mapping of the U.S. and Canada to the same precision as can be obtained using the current file for Alaska.

Other continents, and higher-precision files covering the U.S., may be added later."

vi. Solution to Word Search Puzzle.

```
          P O L Y C O N I C I H P A R G O E R E T S
  L O N G I T U D E E Q U I D I S T A N T     B   C
  A L G O R I T H M     D I V E R G E N T   G I U V
  H               H             A   E     R   R E R
  T         E T O T P M Y S A   L N     A   V C   O
  U L A M B E R T I A C I N O M O N G T   A T S   T
  M R O N         R   N       P P   I   T O   M   A
  I   M           A     G   X   M C   U R     U   T
L Z     A         G C     E     U   R         C O I
  A   F   T       O     Q Q N L C E         L R   O
  R C   R   R   N L   U   U E T R A N S L A T I O N
  O O I   A   I     I     A     I       T H   C   I
  T N   R   C   X N       T     C     I O       N N
  A V     D   T O   J A C O B I A N T G       T   V
  C E       N X A         R       U O       E     E
  R R         I   L       I     D N       G       R
  E G E C I T S L O S     A   E A       R         S
  M E R I D I A N Y       L E L L A R A P         E
            N O I T C E J O R P O B L I Q U E
```

b. "XYNIMAP" — created by David H. Douglas, University of Ottawa; "a comprehensive system for computer cartography and geo-spatial analysis." Preliminary Version. DHDAD@UOTTAWA

"XYNIMAP is a comprehensive system for computer cartography and geo-spatial analysis, that does a lot of things, but not everything, that other packages do. If you give it a chance you will find it does a number of things better than other packages. The diskette contains all manuals and operating instructions. It is meant for PC computers. PC-XT 286, 386, 486.

COMPONENTS

XYNITIZE: An interactive map digitizing system (with a different way of interacting with the user).

BNDRYNET: A program to convert a mass of intersecting lines into a topology to represent the polygons that are visually evident. In other words BNDRYNET is a cartographic spaghetti to polygon converter.

CONSURF: A contour to grid digital elevation model program.

POLYGRID: A polygon to grid converter.

XYBINASC & XYASCBIN: Programs to convert a XYNIMAP *stream feature* file back and forth from compressed binary to readable (therefore editable) ASCII files.

GDEMIDRI: A program to convert a XYNIMAP grid digital elevation model to an IDRISI .img and .doc files.

XYTALLY: A program to read a XYNIMAP stream feature file and produce a printout of various measures: (lengths of lines and areas of regions).

The following are tested workable and distributable programs but I am just not ready to put them out just yet.

XYNIDISP: A comprehensive display system for the PC computer with EGA or VGA graphics adapter cards.

XYNIDRAW: A comprehensive display system for line drawing plotters

VUBLOK: A particularly robust perspective view map program for grid digital elevation models. It produces the traditional fishnet display and shaded relief.

PILLAR: A program to display a geographical distribution by an image of standing vertical pillars on the surface of a perspective view of a base map. The program curves the surface to a realistic projection.

PROCIR: A proportional circle display program."