THE ORIGINS OF CONNECTEDNESS IM KLEINEN

A Dissertation
for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY

John Michael McGrew

1976

This is to certify that the

thesis entitled

THE ORIGINS OF CONNECTEDNESS IM KLEINEN

presented by

John Michael McGrew

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Mathematics

John G. Hocking
Major professor

Date December 12, 1975

THE

Th

the backgr

Theorem.

intermedia

and possib

among the

Ch

by Jordan.

and this i

which math

twentieth

stand the

Ch

curve whic

A countere

attempted

ization do

dimension.

ABSTRACT

THE ORIGINS OF CONNECTEDNESS IM KLEINEN

By

John Michael McGrew

This thesis is an exposition of the history of the background work leading to the Hahn-Mazurkiewicz Theorem. Examples are exhibited illustrating the various intermediate concepts and their range of applicability and possible generalizations. The inter-relationship among the isolated results is demonstrated.

Chapter II explores the definition of curve given by Jordan. Jordan's definition was a very general one, and this is illustrated using the pathological examples which mathematicians of the late nineteenth and early twentieth centuries discovered in their efforts to understand the essence of curve.

Chapter III examines the characterization of curve which Schoenflies gave for subsets of the plane. A counterexample is also discussed, which Brouwer attempted to produce, showing that Schoenflies' characterization does not carry over directly to spaces of higher dimension.

Chapt

Valli, Denjoy

the boundarie

discusses the

to characteri

theory of pri

showing the e

Schoenflies,

also included

Chapt

between 1914

and 1921 are

S" and Hahn'

John Michael McGrew

Chapter IV treats the work of Zoretti, Janiszewski, Nalli, Denjoy and Brouwer on irreducible continua and the boundaries of simply connected regions. Chapter V discusses the independent efforts of Hahn and Mazurkiewicz to characterize Peano spaces. In Chapter VI, Carathéodory's theory of prime ends is presented. Maria Torhorst's work showing the equivalence of the characterizations of Schoenflies, Carathéodory, and Hahn and Mazurkiewicz is also included.

Chapter VII briefly outlines the results obtained between 1914 and 1919. A few selected results from 1920 and 1921 are presented, including Sierpiński's "property S" and Hahn's theorem on the components of open sets.

THE

in par

THE ORIGINS OF CONNECTEDNESS IM KLEINEN

By

John Michael McGrew

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Mathematics

1976

To Mary Kay,
Kay Lynn, Karen, and Catherine

My fi
Divine Influe
during those
wise lost hop
also to my ac
invaluable as
for directin

I ap
with Profess
them contrib
cularly in C

I al
Kelly, J.D.
committee.

I w
Lynn, Karen
appreciated
support thr
University
this work

ACKNOWLEDGMENTS

My first debt of gratitude is to an unmistakable Divine Influence which lifted me up and urged me on during those many trying moments when I might have otherwise lost hope. A special expression of thanks is due also to my adviser, Professor J.G. Hocking for his invaluable assistance in the writing of this thesis and for directing me in my research.

I appreciate, too, the many helpful conversations with Professors P.H. Doyle and D.G. Wright. Both of them contributed materially to the exposition, particularly in Chapter III.

I also wish to thank Professors H.S. Davis, L.M. Kelly, J.D. Schuur, and I. Sinha for serving on my thesis committee.

I want my wife, Mary Kay, and my daughters, Kay Lynn, Karen and Catherine, to know how much I have appreciated their patience, perseverance, sacrifice and support throughout my graduate studies at Michigan State University. Without their cooperation and encouragement this work would have been a near impossibility.

Chapter

TABLE OF CONTENTS

Figure

LIST OF FIGURES

Figure

Figure Page

Writ

once comment

very rare, i

only novel c

of novelty i

may vary con

perience, st

of an event

relative to

position rel

events." [7

made the fol

address as V

the American

"A concept d

from the wav

mathematicia

elements are

mathematical

CHAPTER I

INTRODUCTION

Writing on the history of science, George Sarton once commented: "... creations absolutely de novo are very rare, if they occur at all; most novelties are only novel combinations of old elements, and the degree of novelty is thus a matter of interpretation, which may vary considerably according to the historian's experience, standpoint, or prejudices. ... the determination of an event as the 'first' is not a special affirmation relative to that event, but a general negative proposition relative to an undetermined number of unknown events." [70, p.36] In a similar vein, R.L. Wilder made the following observation during his 1952 retirement address as Vice President and Chairman of Section A of the American Association for the Advancement of Science: "A concept doesn't just pop up full grown 'like Venus from the waves,' although it may seem to, to the individual mathematician who does the conceiving. Usually its elements are lying in what ... we might ... call the mathematical culture stream." [83, p.426]

It v
develop a si
and almost s
from the "ma
illustration
connectednes
individuals
period of al
a very shor
developed fo
during the y
communicatio
to have bee
In fact, 19
edness im k
community a
few years t
to grow at

In
background
Along the w
various int
cability ar
this we dem
isolated co

It very often happens that several individuals develop a significant mathematical idea independently and almost simultaneously, each drawing his inspiration from the "mathematical culture stream." A very good illustration of this process is found in the concept of connectedness im kleinen. It was the work of many individuals in widely scattered locations and over a period of almost a decade, but it came to fruition within a very short span of time and was produced in a well-developed form by two men -- Hahn and Mazurkiewicz -- during the years 1913 and 1914. Probably due to a communication barrier during World War I, they seem not to have been aware of each other's work until nearly 1920. In fact, 1920 may be considered as the point where connectedness im kleinen came to the awareness of the mathematical community as a property with promise, for within the next few years the volume of literature on the subject began to grow at an amazing rate.

In this thesis we explore the mathematical background leading up to the Hahn-Mazurkiewicz Theorem. Along the way, we also exhibit examples illustrating the various intermediate concepts and their range of applicability and possible generalizations. Through all of this we demonstrate the inter-relationship among the isolated concepts.

In C

"curve" give

to be a very

using the pa

the late nin

covered in

curve.

Chap

which Schoe

this chapter

Brouwer atte

characteriza

of higher di

Chap

-- Zoretti,

contemporari

close to def

and Janiszew

Walli, Denjo

of simply co

Chap

and Hahn, in

Peano spaces

prime ends d

work showing

In Chapter II we explore the definition of "curve" given by Jordan. Jordan's definition turned out to be a very general one indeed, and we illustrate this using the pathological examples which mathematicians of the late nineteenth and early twentieth centuries discovered in their efforts to understand the essence of curve.

Chapter III examines the characterization of curve which Schoenflies gave for subsets of the plane. In this chapter we also discuss a counterexample, which Brouwer attempted to produce, showing that Schoenflies' characterization does not carry over directly to spaces of higher dimension.

Chapter IV treats the work of several mathematicians -- Zoretti, Janiszewski, Nalli, Denjoy and Brouwer -- all contemporaries of Hahn and Mazurkiewicz -- who came very close to defining connectedness im kleinen. Zoretti and Janiszewski were studying irreducible continua, while Nalli, Denjoy, and Brouwer were each examining boundaries of simply connected domains.

Chapter V discusses the efforts of Mazurkiewicz and Hahn, independently of each other, to characterize Peano spaces. In Chapter VI we present the theory of prime ends developed by Carathéodory. Maria Torhorst's work showing the logical equivalence of the concepts involved in t

curves due to

Mazurkiewicz

In Cl

obtained betw

results from

"property S

open sets.

There

works giving

Kleinen (loc

for example,

of Point Set

or R.L. Wild

and clarify

In keeping w

its developm

Unle

author's own

are cited fr

ular form

As a genera

in the text

interest or

involved in the three characterizations of simple closed curves due to Schoenflies, Carathéodory, and Hahn and Mazurkiewicz is also included in this chapter.

In Chapter VII we briefly outline the results obtained between 1914 and 1919 and present a few selected results from 1920 and 1921, including Sierpiński's "property S" and Hahn's theorem on the components of open sets.

There are already in existence many technical works giving a development of the theory of connected im kleinen (locally connected) spaces in some detail (see, for example, Kuratowski [46], R.L. Moore's Foundations of Point Set Theory, G.T. Whyburn's, Analytic Topology, or R.L. Wilder [82].) Our goal has been to investigate and clarify only the origins of connectedness im kleinen. In keeping with this goal, we have not attempted to trace its development beyond 1921.

Unless otherwise stated all proofs are the author's own. Though the results of other mathematicians are cited frequently, the responsibility for the particular form used here rests solely with the present author. As a general rule, proofs of such results are not repeated in the text unless the method used is either of historical interest or is referred to in our exposition.

We a

theory, and

operations

consideratio

complement o

context. We

ground a fir

material pre

Young [29].

Not

Therefore, t

understands

We make the

this thesis

space X i

The ball ne

denoted $N($

neighborhoo

p.61]), is

subset N c

N itself n

variation i

significant

We assume that the reader has a knowledge of set theory, and we will use the customary symbols for set operations ($\in, \subset, \cup, \cap, -$). In addition, for any set under consideration, we use the notation $\mathcal{C}A$ to denote the complement of A, the universe being understood from context. We also assume as part of the reader's background a first year course in topology equivalent to the material presented in Dugundji [19], or in Hocking and Young [29].

Not all authors agree on definitions and notation. Therefore, the reader is cautioned to always make sure he understands which ones are intended. With this in mind, we make the following clarification. As used throughout this thesis, a <u>neighborhood</u> of a point p in a topological space X is any open subset of X which contains p. The ball neighborhood centered at p with radius r is denoted N(p,r). Another very prevalent definition of neighborhood, used, for example, by Kuratowski (see [45, p.61]), is this one: a neighborhood of $p \in X$ is any subset $N \subset X$ with p in the interior of N. Note that N itself need not be open. As we shall see, this slight variation in the definition of neighborhood makes a significant difference in another definition.

We

of the thes

space is co

for every r

V of x c

of V ther

and x co

kleinen if

points.

A

points and

topologis

kleinen a

This exa

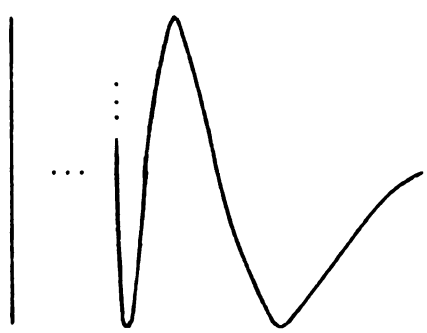
not conn

be conn

shown b

We turn our attention now to the central concept of the thesis -- connectedness im kleinen. A topological space is <u>connected im kleinen at the point</u> $x \in X$ if, for every neighborhood U of x, there is a neighborhood V of x contained in U such that for every point p of V there is a connected subset C of U with p and x contained in C. X is said to be <u>connected im kleinen</u> if it is connected im kleinen at each of its points.

A set may be connected im kleinen at some of its points and not connected im kleinen at others. The topologist's sine wave, for example, is not connected im kleinen at any of its points along the y-axis.



This example illustrates that a set may be connected but not connected im kleinen. By the same token, a set may be connected im kleinen without being connected, as is shown by considering any discrete space.

Perhaps

of the next two

numbers between

be the radius

with the posit

connected im k

On t

Kleinen at

constructed

real line v

the relati

set is cor

A

of connec

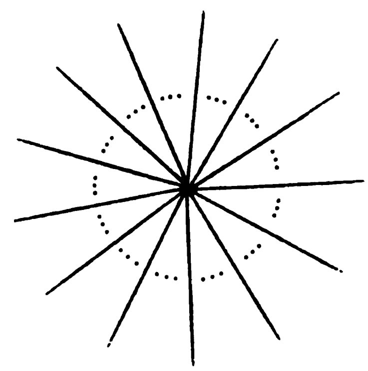
characte

which ca

unit in

what is

Perhaps a little more surprising are the properties of the next two examples. Let T be the set of rational numbers between 0 and 2π. For each $\theta \in T$, let r_θ be the radius of the unit circle forming an angle of θ with the positive x-axis. Let $R = \bigcup_{\theta \in T} r_\theta$. Then R is connected im kleinen at the origin but at no other point.



On the other hand, a set may be connected im kleinen at every point except one. Such a set can be constructed by taking the union of the origin on the real line with the points of the sequence $\{1/n\}_{n=1}^{\infty}$. With the relative topology inherited from the real line, this set is connected im kleinen at every point except the origin.

As we shall point out again later, the concept of connectedness im kleinen was originally developed to characterize "continuous curve" -- that is, any point set which can be expressed as a continuous image of the closed unit interval. This characterization is embodied in what is now called the Hahn-Mazurkiewicz Theorem, namely:

A point set is

connected, con

Chapter V).

Since

but disappeare

a few words ab

connectedness.

is given here

if, for every

connected nei

said to be lo

at each of it

The d

be used for e

of neighborho

neighborhood

is precisely

is connected

of x with

that for each

$_p \subset U$ with

neighborhood

reverse impl

neighborhood

connected su

A point set is a continuous curve iff it is compact, connected, connected im kleinen and metrizable (see Chapter V).

Since the term "connectedness im kleinen" has all but disappeared from the current literature we include a few words about its more modern counterpart, "local connectedness." For the sake of comparison, the definition is given here. X is <u>locally connected at the point</u> $x \in X$ if, for every neighborhood U of x, there exists a connected neighborhood V of x contained in U. X is said to be <u>locally connected</u> if it is locally connected at each of its points.

The definition of local connectedness can actually be used for either concept depending on which definition of neighborhood is intended. With the definition of neighborhood as used by Kuratowski, local connectedness is precisely connectedness im kleinen. That is, if X is connected im kleinen at x and U is a neighborhood of x with V the corresponding subneighborhood such that for each point $p \in V$ there exists a connected set $C_p \subset U$ with $p, x \in C_p$, then $C = \bigcup_{p \in V} C_p$ is a connected neighborhood of x in the sense of Kuratowski. The reverse implication follows as easily, since, for each neighborhood U of a point $x \in X$, the existence of a connected subset C of U with x in $\mathrm{Int}(C)$ implies that there is

this set C i

and any point

If, ho

set, then the

following set

illustrates t

p

X is conne

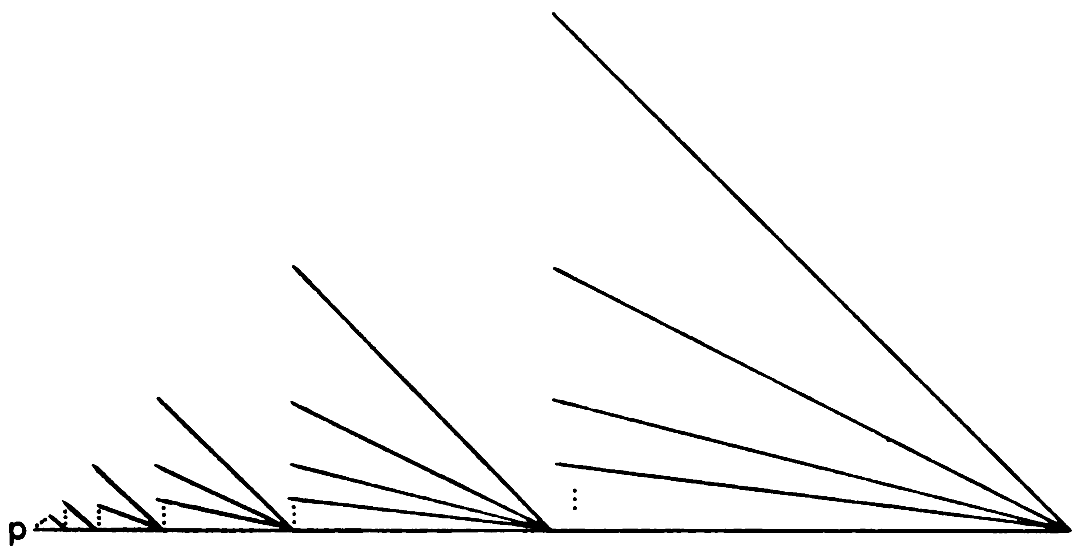
connected a

shows, the

equivalent.

that there is an open neighborhood $V \subset C$ with $x \in V$. This set C is a connected subset of U containing x and any point p of V.

If, however, by a neighborhood we mean an open set, then the two concepts are not equivalent. The following set K (based on an example in [29, p.113]) illustrates this distinction:



K is connected im kleinen at p, but it is not locally connected at p. However, as the following theorem shows, the two definitions are, in a certain sense, equivalent.

Theore

connected im l

Proof

That

every point c

point of X.

We c

from topolog

thesis.

The

bounded sub

for a proof

In

the follow

T

and let

X onto

[26, p.7

4.1.

<u>Theorem 1.1</u>. X is locally connected iff it is connected im kleinen.

<u>Proof</u>: See [29, p.114].

That is, if one of the properties holds at <u>every</u> point of X then the other also holds at every point of X.

We conclude this chapter with some basic results from topology and analysis which we use implicitly in the thesis.

<u>Theorem 1.2</u>. (Heine-Borel). Every closed and bounded subset of E^n is compact. (See [69, p.42] for a proof of the case $n = 1$.)

In discussing the Brouwer sphere (Chapter III) the following fact is assumed:

<u>Theorem 1.3</u>. Let X and Y be subsets of E^n, and let $f:X \to Y$ be a one-to-one continuous function from X onto Y. If X is compact then f is a homeomorphism. [26, p.76]

The next result is used in the proof of Theorem 4.1.

Theor

compact iff i

The p

immediate con

sequence.

Theor

numbers conve

subsequence.

<u>Theorem 1.4</u>. A subset of a compact space is compact iff it is closed. [19, p.224]

The proof of Lemma 4.9 uses the following immediate consequence of the definition of a convergent sequence.

<u>Theorem 1.5</u>. Every sequence of positive real numbers converging to zero has a monotone decreasing subsequence.

Babylon I

time of A

sections

order cu

"cissoid'

came unde

curves is

of "curve

mathemati

an inters

point.

W

applicati

either as

means of

---

1) $x = 2$

2) $x = 2$

3) $x = \frac{3}{1-}$

CHAPTER II

WHAT IS A CURVE?

The mathematicians of ancient Egypt and of Babylon knew some basic properties of the circle. By the time of Archimedes much was known about the conic sections and several other classes of curves. Higher order curves such as the "witch" of Agnesi[(1)], the "cissoid" of Diocles[(2)], and the "folium" of Descartes[(3)] came under study long ago. In short, the study of curves is as old as mathematics itself. The definition of "curve" must have seemed obvious to these early mathematicians. A curve was defined geometrically as an intersection of surfaces or as the locus of a moving point.

With the development of analysis and its applications to mechanics, a curve came to be considered either as the graph of a function or as being defined by means of parametric equations. A continuity condition

---

1) $x = 2r\cdot\tan(\theta)$, $y = 2r\cdot\cos^2(\theta)$, $y(x^2+4r^2) = 8r^3$.

2) $x = 2r\cdot\sin^2(\theta)$, $y = 2r\cdot\sin^2(\theta)\cdot\tan(\theta)$, or $x^3 = y^2(2r-x)$.

3) $x = \dfrac{3t}{1+t^3}$, $y = \dfrac{3t^2}{1+t^3}$, $x^3 + y^3 = 3xy$.

was assumed, (
conditions as
of the ninete
"curve" was s
functions was
Georg Cantor
establish a c
of a line anc
take a decima
the point of
$x = 0.a_1a_3a_5$
[62] (in 187
was necessar
Camille Jord
very genera
image of a
K. Weierstr
real variab
The example
and it seem
Only three
was shown
strange "c

On
was the di
Guiseppe

was assumed, of course, and usually some differentiability conditions as well. However, during the final decades of the nineteenth century the complacent acceptance of "curve" was shattered. The discovery of pathological functions was the cause. One such was described by Georg Cantor in 1878 [11]. Cantor showed that one can establish a one-to-one correspondence between the points of a line and those of a surface. (The basic idea is to take a decimal number $t = 0.a_1a_2a_3a_4\ldots$ and map it onto the point of the square with coordinates (x,y), where $x = 0.a_1a_3a_5\ldots$ and $y = 0.a_2a_4a_6\cdots$.) But Eugen Netto [62] (in 1879) and others showed that such a correspondence was necessarily discontinuous. This may have encouraged Camille Jordan in his Cours d'analyse of 1887 to give a very general definition, namely, a curve is a continuous image of a line segment. It had been shown in 1861 by K. Weierstrass that there are continuous functions of a real variable that possess no tangent line at any point. The examples known at that time were one-dimensional, and it seems that they caused Jordan little concern. Only three years after he first published it, however, it was shown that Jordan's definition provides some very strange "curves" indeed.

One of the most unsettling revelations of the time was the discovery of "space-filling" curves. In 1890, Guiseppe Peano published a description of two continuous functions -- c

square and the

interval onto

construction w

readily seen,

Cantor's one-t

In his

number in the

in powers of

where $a_i \in \{$

defined a seq

digits a_i,

Next, to eac

correspond t

functions -- one from the unit interval onto the unit square and the other, a similar one, from the unit interval onto the unit cube [66]. Peano's original construction was purely arithmetic. In fact as will be readily seen, it was merely a clever modification of Cantor's one-to-one correspondence.

In his construction, Peano represented each number in the interval $0 \leq t \leq 1$ as an infinite series in powers of $1/3$. That is,

$$t = \frac{a_1}{3^1} + \frac{a_2}{3^2} + \ldots + \frac{a_n}{3^n} + \ldots$$

where $a_i \in \{0,1,2\}$ for each $i = 1,2,\cdots$. He then defined a sequence of operators, $k, k^2, k^3, \ldots,$ on the digits a_i, by

$$\begin{aligned} ka_i &= 2 - a_i, \\ k^2 a_i &= k(ka_i), \\ &\vdots \\ k^n a_i &= k(k^{n-1} a_i) \\ &\ldots \end{aligned}$$

Next, to each t, a point of the unit interval, he let correspond the point (x,y) of the Cartesian plane, where

$$x = \frac{b_1}{3} + \frac{b_2}{3^2} + \ldots + \frac{b_n}{3^n} + \ldots$$

$$y = \frac{c_1}{3} + \frac{c_2}{3^2} + \ldots + \frac{c_n}{3^n} + \ldots$$

and the digits

$b_1 = a$

$b_3 = k$

$b_n = k^{a_2+a}$

Although thes
Peano demonst
well-defined
tinuous funct

Peano
paper, presum
arrived at a
of a diagram
was led to t
such diagram
paper leaves
for what has
a set of equ
E. Cesaro [

A y
Hilbert [28
example of
simplificat
most often

and the digits b_i and c_i are given by

$$b_1 = a_1, \; c_1 = k^{a_1}(a_2), \; b_2 = k^{a_2}(a_3), \; c_2 = k^{a_1+a_3}(a_4),$$
$$b_3 = k^{a_2+a_4}(a_5), \ldots,$$
$$b_n = k^{a_2+a_4+\ldots+a_{2n-2}}(a_{2n-1}), \; c_n = k^{a_1+a_3+\ldots+a_{2n-1}}(a_{2n}), \ldots.$$

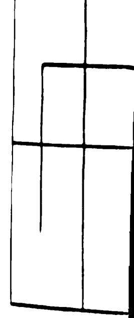
Although these representations are not necessarily unique, Peano demonstrates that the correspondence is nonetheless well-defined and the coordinates x and y are continuous functions of t.

Peano included no geometric illustrations in his paper, presumably so that no one would think he had arrived at a false proof through the misinterpretation of a diagram. One cannot but believe, however, that Peano was led to the discovery of his curve by studying just such diagrammatic representations. In any event, his paper leaves the reader with very little intuitive feeling for what has happened! It was not until 1897 that even a set of equations for Peano's curve was published (see E. Cesaro [15]).

A year after Peano published his results, David Hilbert [28] provided a somewhat simpler and geometric example of a space-filling curve. It is Hilbert's simplification (or some modification of it) which we most often call the Peano curve.

We giv

three stages c



(a)

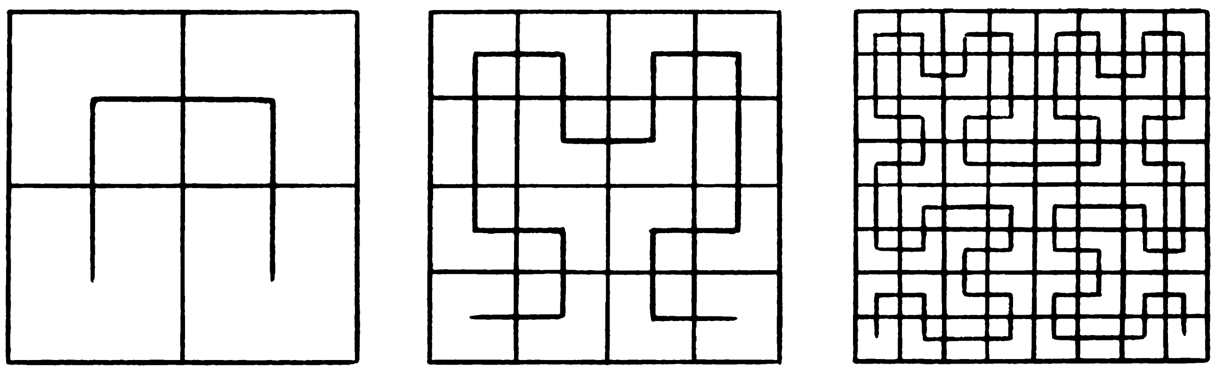
In

[53] each c

original c

(a)

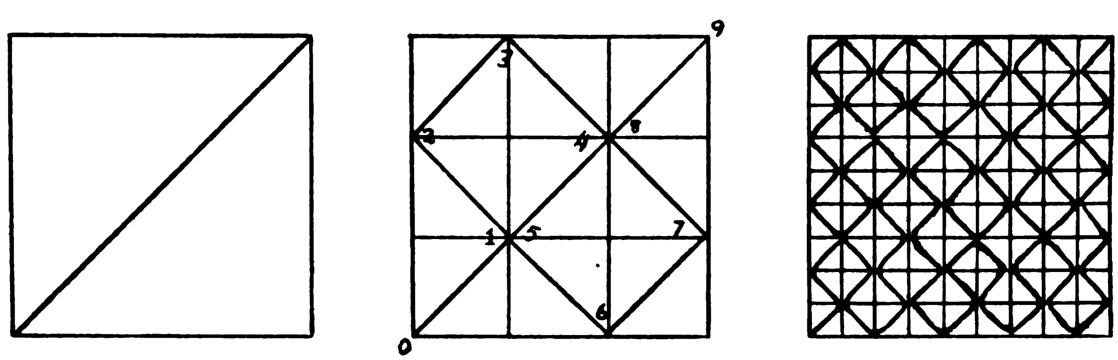
We give here a pictorial description of the first three stages of Hilbert's construction:



(a) (b) (c)

Figure 1

In 1900, Arthur Schoenflies [71] and E.H. Moore [53] each gave a geometric representation of Peano's original construction as follows:



(a) (b) (c)

Figure 2

Mc

coordinate

curve (see

functions

of Hilbert

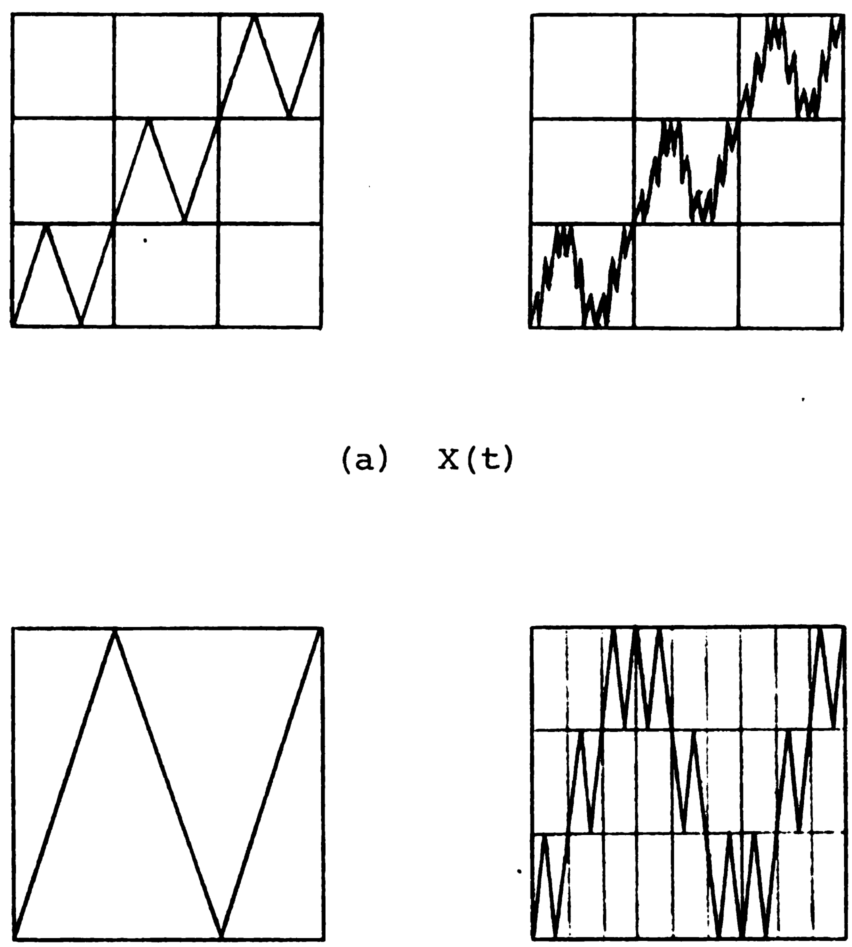
curve is a

$t = 1$ ar

Figure 4

construct

Moore gave a proof of the fact that the coordinate functions X(t) and Y(t) of the Peano curve (see Figure 3) are continuous, nowhere differentiable functions of t. He also described a simple modification of Hilbert's curve such that the resulting continuous curve is a closed curve, that is, such that $t = 0$ and $t = 1$ are both mapped to the same point of the square. Figure 4 indicates the first three stages of this modified construction.



(a) X(t)

(b) Y(t)

Figure 3

It

Peano's cur

In fact eac

or four poi

unit squar

at the ori

Hilbert's

square who

exactly fo

points of

one) of t

the squar

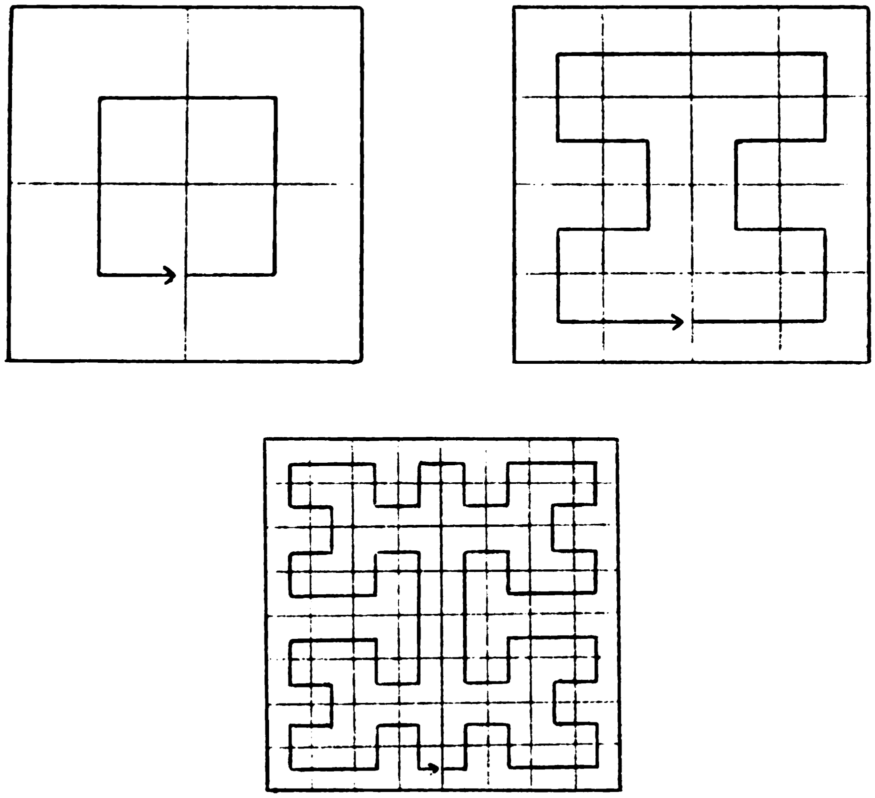


Figure 4

It will be noted that neither Hilbert's curve nor Peano's curve is a one-to-one image of the unit interval. In fact each point of the square is the image of one, two, or four points of the unit interval. If we consider the unit square as having its lower left-hand corner located at the origin of a Cartesian coordinate system, then for Hilbert's curve, all those points in the interior of the square whose coordinates are of the form $(\frac{i}{2^k}, \frac{j}{2^k})$ have exactly four pre-images in the unit interval; those points of the interior with only one coordinate (either one) of the form $\frac{i}{2^k}$ and those points of the boundary of the square with both coordinates of that form (with the exception of t

images; and th

coordinate of

and those bou

the form $\frac{i}{2^k}$

Peano's curve

description.)

Hilbe

by a suitable

achieve a map

of the square

interval. Ha

Abbildung ein

1913 [20], de

curve (see Fi

Hilbert's cur

paper that th

one, was gene

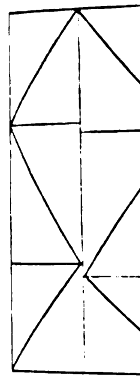
reference to

be determined

the literatu

exception of the corner points) have exactly two pre-images; and those points of the square with neither coordinate of the form $\frac{i}{2^k}$, plus the four corner points and those boundary points with only one coordinate of the form $\frac{i}{2^k}$ have exactly one pre-image. (Note: for Peano's curve, we need only replace 2 by 3 in the above description.)

Hilbert makes the observation in his paper that by a suitable change in the partition lines one can achieve a mapping such that the pre-image of any point of the square contains at most three points of the unit interval. Hans Hahn, in his paper entitled Über die Abbildung einer Strecke auf ein Quadrat, published in 1913 [20], describes such a modification for Peano's curve (see Figure 5). He was obviously familiar with Hilbert's curve, and one gets the impression from his paper that the construction we give here, or a similar one, was generally known, though Hahn gives no specific reference to where it might be found, and, as far as can be determined, such a construction does not appear in the literature.



In co

carries out t

necessary cha

more carefull

this can be d

parentheses i

is reversed.)

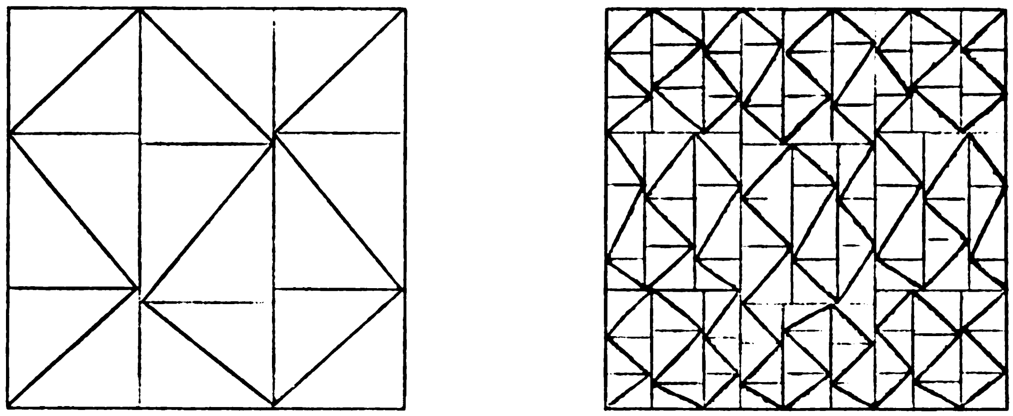


Figure 5

In contrast to the simplicity with which Hahn carries out the modification for Peano's curve, the necessary changes in Hilbert's construction must be made more carefully. We include here a description of how this can be done. (In Figure 6, the lettering in parentheses is assumed if the direction of the "horseshoe" is reversed.)

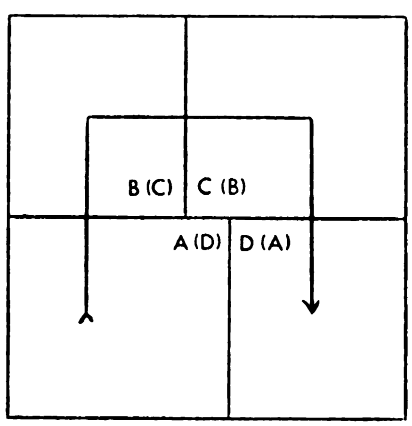


Figure 6

The

Hilbert's ex

at most thre

point.

By a

rectangles w

the rectangl

construction

group of si

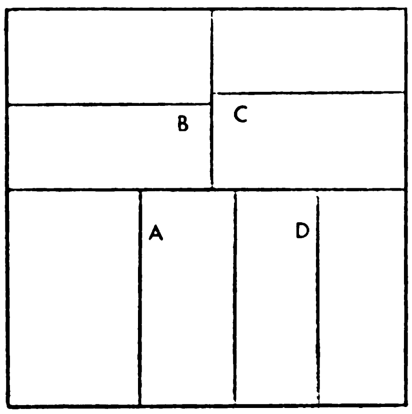
division of

back in the

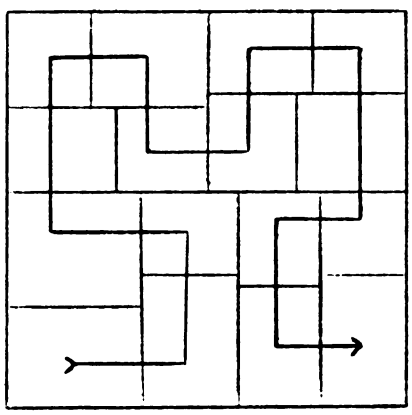
block of fo

The basic idea is to offset the division lines in Hilbert's example so that at any stage of the construction at most three regions of the square meet at any given point.

By a block of four we shall mean any group of four rectangles which together form a subdivision of one of the rectangles resulting from the previous stage of construction. By a block of sixteen we shall mean any group of sixteen rectangles which together form a subdivision of one of the rectangles occurring two stages back in the construction. (Examples: Figure 6 is a block of four and Figure 7(b) is a block of sixteen.)



(a)



(b)

Figure 7

Figure

be used throu

the regions

refer to the

relative to t

the curve. T

$c \in f^{-1}(C)$ a

We pr

positive real

Stage

parts by cons

the midpoints

units of the

two non-inte

the upper ha

other separa

label the re

as in Figure

Supp

been complet

resulting co

labeled B

a division

and is locat

boundary li

Figure 6 illustrates the basic configuration to be used throughout our construction. The labeling of the regions A,B,C, and D given there will be used to refer to the position of each region in a block of four relative to the directed "horseshoe" configuration of the curve. That is, if $a \in f^{-1}(A)$, $b \in f^{-1}(B)$, $c \in f^{-1}(C)$ and $d \in f^{-1}(D)$, then $a < b < c < d$.

We proceed by induction. Let p be a fixed positive real number less than $\frac{1}{2}$.

<u>Stage 1</u>: We divide the square into two equal parts by constructing a horizontal division line between the midpoints of the two vertical sides. Within p units of the midpoint of this division line, we construct two non-intersecting vertical lines -- one separating the upper half of the square into two parts and the other separating the lower half of the square. We now label the regions A,B,C,D, and construct the horseshoe as in Figure 6.

Suppose now that the first $k - 1$ stages have been completed. Consider any block of four in the resulting configuration. We divide each of the regions labeled B and C into two non-empty parts by inserting a division line in each which meets their common boundary and is located within $p \cdot h$ units of the midpoint of that boundary line, where h is the length of the boundary line. Now div

is parallel to

and located wi

segment $\overline{rs}$

of the segmen

These and all

so that in th

four regions

We no

more division

specify how t

being complet

line already

division line

Each of these

within p·h'

line, where

Now label th

and form the

curve shown

stage 3 of t

line. Now divide region A such that the division line is parallel to the common boundary between A and D and located within $p \cdot h_{\overline{rs}}$ units of the midpoint of the segment $\overline{rs}$ (see Figure 7(a)), where $h_{\overline{rs}}$ is the length of the segment $\overline{rs}$. Similarly divide the region D. These and all other division lines will be constructed so that in the resulting subdivision of the square, no four regions will meet in a single point.

We now complete the subdivision by adding eight more division lines to form a block of sixteen. We specify how this is done for region A only, the rest being completed in the same fashion. On the division line already constructed in region A, erect two more division lines to form, within A, a block of four. Each of these two new division lines is to be located within $p \cdot h'$ units of the midpoint of the former division line, where h' is the length of that division line. Now label this new block of sixteen as shown in Figure 7(b) and form the horseshoe of the previous stage into the curve shown in Figure 7(b). Figure 8 illustrates stage 3 of the construction.

Noti

and the Pea

mesh of the

mesh of any

of a diagona

One naturall

presented ab

for introduc

At the end o

maximum poss

rectangles i

upper bound

of the secon

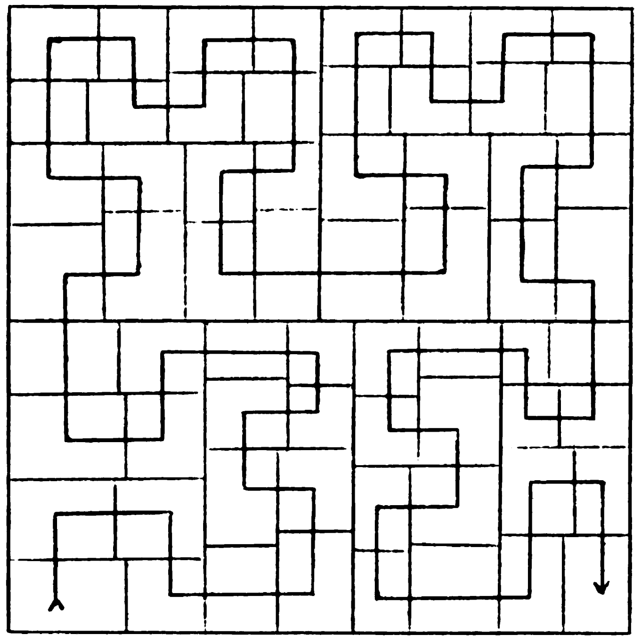


Figure 8

Notice that the success of the Hilbert curve (and the Peano curve itself) depends on the fact that the mesh of the subdivisions tends to zero. (In fact the mesh of any of the subdivisions is precisely the length of a diagonal of any one of the squares of the subdivision.) One naturally asks, "Is this true of the construction presented above?" Yes! This is precisely the reason for introducing the constant p into the construction. At the end of the first stage of the construction, the maximum possible length of the sides of the resulting rectangles is $\frac{1}{2} + p$. (This is actually an attained upper bound for the lengths of the sides.) At the end of the second stage, the lengths are all less than

$(\frac{1}{2} + p)^2$,

are all 1

tends to

as we hav

T

line of r

than two

consists

of Peano

twenty ye

curves be

described

S

approach

proves

I

function

even and

(1)

for all

(2)

for $0 \leq$

$(\frac{1}{2} + p)^2$, and at the end of the kth stage the lengths are all less than $(\frac{1}{2} + p)^k$, which tends to zero as k tends to infinity provided p is strictly less than $\frac{1}{2}$, as we have specified.

The announcement of Peano's curve initiated a line of research which continued sporadically for more than two decades. The first phase of this research consists only of a series of elaborations of the examples of Peano and Hilbert. This phase lasted for nearly twenty years. Then in 1912 new examples of space-filling curves began to appear, the first of which is the one described below due to Waclaw Sierpiński.

Sierpiński [77] takes a slightly different approach to defining a space-filling curve. He first proves

Theorem 2.1. There exists a unique f, a function of the real variable t, which is bounded and even and satisfies the functional equation:

(1) $$f(t) + f(t + \tfrac{1}{2}) = 0,$$

for all real numbers t, and the equation:

(2) $$2f(t/2) + f(t + 1/8) = 1,$$

for $0 \leq t \leq 1$.

T

auxiliary

of period

(2), Sie

i

f(

ii

ii, one

(3) f

The proof of this theorem utilizes the two auxiliary functions $\theta(t)$ and $\tau(t)$, both periodic of period 1, defined by

$$\begin{cases} \theta(t) = 1\ , & \tau(t) = 1/8 + 4t \quad \text{on} \quad 0 \le t < \frac{1}{4} \\ \theta(t) = -1, & \tau(t) = 1/8 - 4t \quad \text{on} \quad \frac{1}{4} \le t < \frac{1}{2} \\ \theta(t) = -1, & \tau(t) = 1/8 + 4t \quad \text{on} \quad \frac{1}{2} \le t < \frac{3}{4} \\ \theta(t) = 1\ , & \tau(t) = 1/8 - 4t \quad \text{on} \quad \frac{3}{4} \le t < 1. \end{cases}$$

With a little manipulation of equations (1) and (2), Sierpiński establishes the following facts:

i. f is periodic of period 1;

ii.
$$f(t) = \begin{cases} \frac{1}{2} - \frac{1}{2}\, f(1/8 + 4t) & \text{on} \quad 0 \le t \le \frac{1}{4} \\ -\frac{1}{2} + \frac{1}{2}\, f(1/8 - 4t) & \text{on} \quad \frac{1}{4} \le t \le \frac{1}{2} \\ -\frac{1}{2} + \frac{1}{2}\, f(1/8 + 4t) & \text{on} \quad \frac{1}{2} \le t \le \frac{3}{4} \\ \frac{1}{2} - \frac{1}{2}\, f(1/8 + 4t) & \text{on} \quad \frac{3}{4} < t \le 1; \end{cases}$$

iii. for all real numbers, t,

$$f(t) = \frac{\theta(t)}{2}\,[1-f(\tau(t))].$$

By successive replacement of t by $\tau(t)$ in iii, one obtains

$$(3) \qquad f(t) = \frac{\theta(t)}{2} - \frac{\theta(t)\cdot\theta(\tau(t))}{2^2} + \frac{\theta(t)\cdot\theta(\tau(t))\cdot\theta(\tau^2(t))}{2^3} - \cdots + \frac{(-1)^n\cdot\theta(t)\cdot\theta(\tau(t))\cdot\ldots\cdot\theta(\tau^n(t))}{2^{n+1}}\cdot f(\tau^n(t)).$$

Since f is

right hand s

and therefor

(4) $\frac{\theta(t}{2}$

converges to

establishes

theorem. Th

of $f(t)$ to

verges, then

of Theorem 1

Sier

<u>Theo</u>

(5)

define a con

(We

involves onl

The

Figure 9 sho

Sierpiński's

Since f is bounded, the last term in the sum on the right hand side of (3) tends to zero with increasing n, and therefore the infinite series

$$(4) \qquad \frac{\theta(t)}{2} - \frac{\theta(t)\cdot\theta(\tau(t))}{2^2} + \frac{\theta(t)\cdot\theta(\tau(t))\cdot\theta(\tau^2(t))}{2^3} - \cdots$$

converges to $f(t)$ for all real numbers t. This establishes the uniqueness of f as asserted in the theorem. The existence follows by defining the value of $f(t)$ to be the value to which the series (4) converges, then demonstrating that all of the conditions of Theorem 1 are satisfied by this function.

Sierpiński next asserts

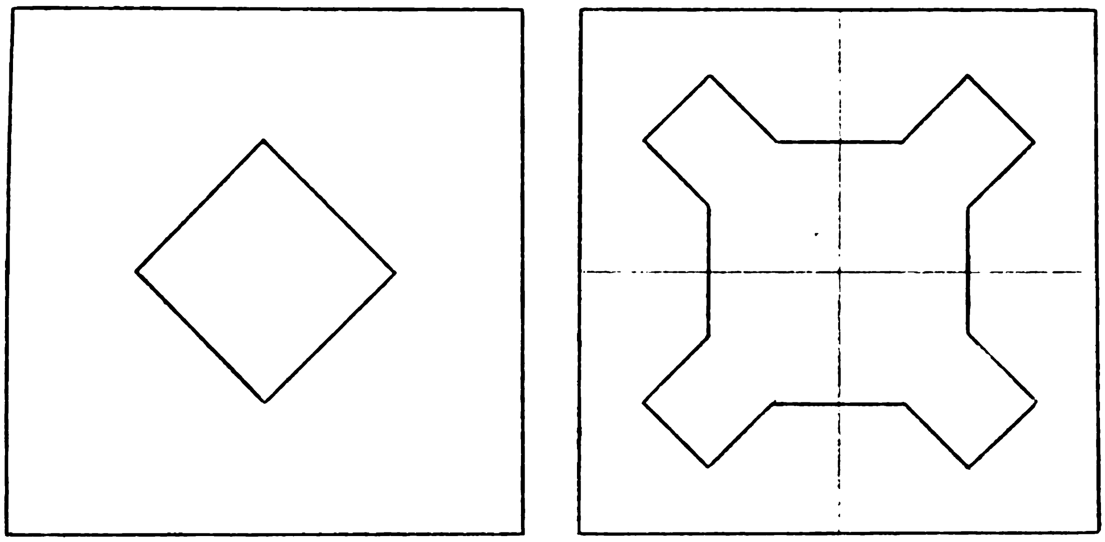
<u>Theorem 2.2</u>. The equations

$$(5) \qquad \left.\begin{aligned} x &= f(t) \\ y &= f(t - \tfrac{1}{4}) \end{aligned}\right\} \; 0 \leq t \leq 1$$

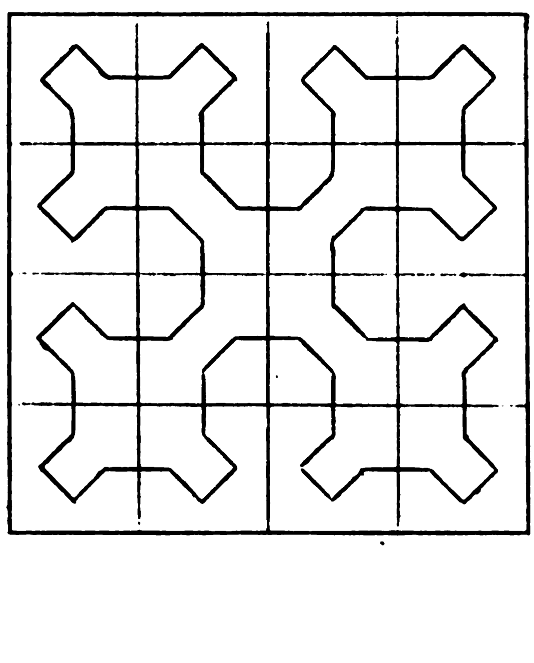
define a continuous space-filling curve.

(We omit any indication of the proof since it involves only standard arguments.)

The curve given by (5) is in fact a closed curve. Figure 9 shows a simple geometric interpretation of Sierpiński's curve.



(a)



(b)

Figure 9

G. P

illing a ri

riangle has

nit interva

epresentati

Poly

are of unequ

of the trian

is the verte

longer leg

is then drop

the base of

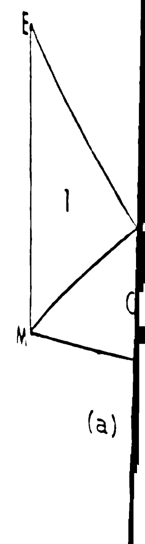
triangle OM

MP_0O. Label

triangles M

triangle MP

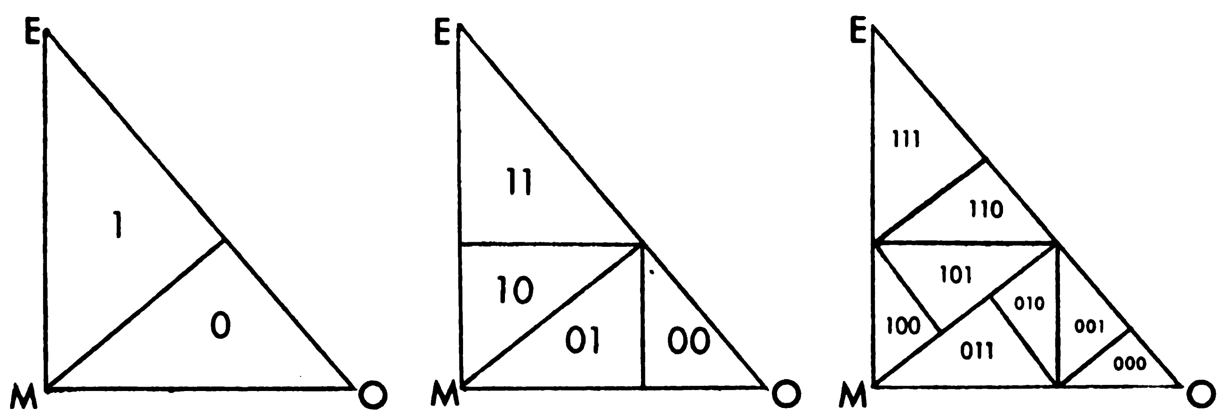
new triangle



(a)

G. Polya [67] gives a construction of a curve filling a right triangle, such that each point of the triangle has at most three distinct pre-images in the unit interval. His construction uses the binary representation of the numbers in the unit interval.

Polya starts with a right triangle whose legs are of unequal length. (See Figure 10) The vertices of the triangle are labeled O, M, and E, where M is the vertex at the right angle, E determines the longer leg and O the shorter leg. A perpendicular is then dropped from M to the hypotenuse OE. Denote the base of this perpendicular by P_o. MP_o divides triangle OME into two smaller triangles, MP_oE and MP_oO. Label the smaller of these two triangles, triangles MP_oO, with a 0, and label the larger triangle MP_oE, with a 1. Note that each of the two new triangles is similar to triangle OME.



(a) (b) (c)

Figure 10

P_0P_1 ar

label ea

11, acc

of the s

division

0 or 1

each tri

subtrian

perpendi

stages c

pondence

triangle

continuo

the tria

points o

of the t

the case

But Poly

that, fo

integer,

The proo

interrel

From point P_0 construct two more perpendiculars, P_0P_1 and P_0P_1', to ME and MO repsectively, and label each of the four new triangles 00 or 01, 10 or 11, according as it is the smaller or larger subtriangle of the smaller or larger triangle of the previous subdivision. This process is continued indefinitely and a 0 or 1 added to the binary sequence associated with each triangle according as it is the smaller or larger subtriangle created by the introduction of a new perpendicular. Figure 10 illustrates the first three stages of this process.

The limit of this construction gives a correspondence between binary sequences and points of the triangle, which, as Polya shows, gives us a well-defined continuous surjection from the closed unit interval to the triangle OME.

It is still not evident that no more than three points of the unit interval are mapped to the same point of the triangle. In fact, it is not true that such is the case for all right triangles with unequal legs. But Polya demonstrates that if the triangle is chosen so that, for α = the angle MOE, $1/\cos^2 \alpha$ is not an algebraic integer, then the desired condition does indeed hold. The proof of this fact is very elegant and shows an interrelationship between geometry and algebraic number theory. W

which Poly

observatio

Le

to which a

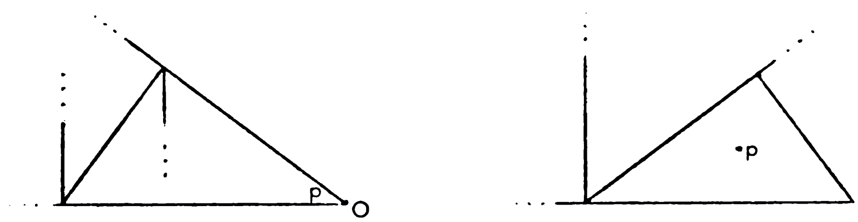
any stage

all the po

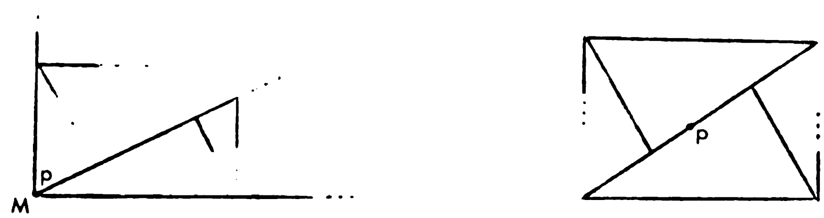
1,2,4,5,

theory. We will present it in essentially the form in which Polya originally gave it. But first we make one observation.

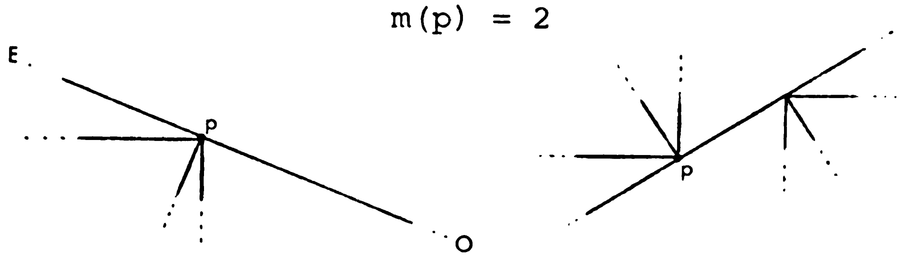
Let m(p) denote the maximum number of subtriangles to which a point p of the triangle OME can belong at any stage of the construction. Figure 11 illustrates all the possible values m(p) can assume -- namely, 1,2,4,5, and 8.



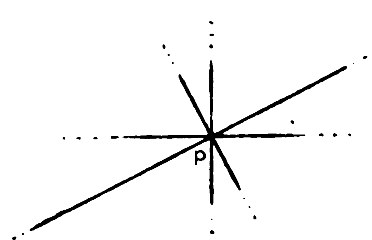
m(p) = 1



m(p) = 2



m(p) = 4 m(p) = 5



m(p) = 8

Figure 11

N

We will c

lines whi

from 0.

E

M

only o

distan

where

not

point

their

and

Now let ℓ be the length of the hypotenuse, OE. We will consider only those points of the subdivision lines which fall on OE and determine their distances from O.

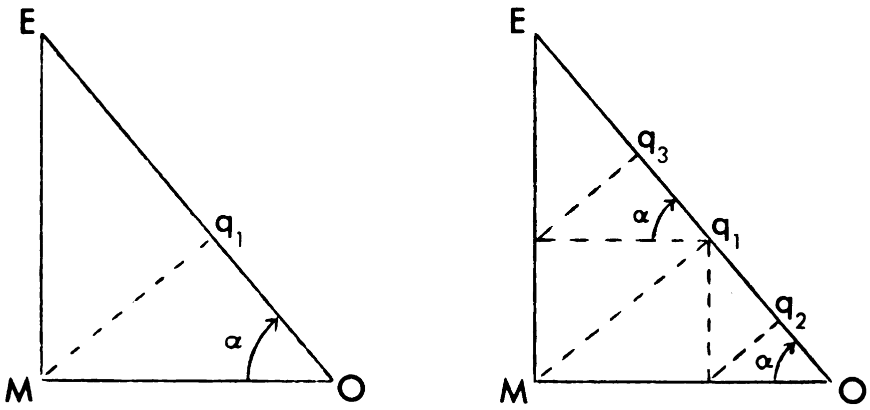


Figure 12

At the first stage of construction (see Figure 12), only one such point, q_1 (= P_o), is introduced. Its distance from O is

$$\ell \cdot \cos^2 \alpha = \ell \cdot R_1(\cos^2 \alpha),$$

where $R_1(\cos^2 \alpha)$ denotes "R_1 is a function of $\cos^2 \alpha$," not "R_1 times $\cos^2 \alpha$."

In the second and third stages two more such points, q_2 and q_3, are introduced along OE, and their distances are given by

$$(\ell \cdot \cos^2 \alpha)\cos^2 \alpha = \ell \cdot \cos^4 \alpha = \ell \cdot R_2(\cos^2 \alpha)$$

and

$(\ell-\ell\cdot\cos^2$

respectively

In t

polynomial f

$R_1($

and a seque

ℓR_1

representin

ponding to

hypotenuse

Bec

all of the

analysis ap

is a point

unit interv

subtriangle

two triangl

all eight c

diagonal be

and ν,

that is,

(*)

$$(\ell - \ell \cdot \cos^2 \alpha)\cos^2 \alpha + \ell \cdot \cos^2 \alpha = \ell \cdot (2 \cos^2 \alpha - \cos^4 \alpha)$$
$$= \ell \cdot R_3(\cos^2 \alpha)$$

respectively.

In this way Polya generates a sequence of polynomial functions in $\cos^2 \alpha$,

$$R_1(\cos^2 \alpha),\ R_2(\cos^2 \alpha), \ldots, R_n(\cos^2 \alpha), \ldots,$$

and a sequence

$$\ell R_1(\cos^2 \alpha),\ \ell R_2(c9s^2 \alpha), \ldots, \ell R_n(\cos^2 \alpha), \ldots,$$

representing the distances (as measured from O) corresponding to the points of the subdivisions lying on the hypotenuse OE.

Because of the similarity to triangle OME of all of the subtriangles of each subdivision, the same analysis applies to each of these subtriangles. If there is a point of the triangle having four pre-images in the unit interval, then at some stage it must belong to eight subtriangles and will lie on the common hypotenuse of two triangles which together form a rectangle containing all eight of these subtriangles. Let the length of this diagonal be ℓ'. Then for some positive integers μ and ν,

$$\ell' R_\mu(\cos^2 \alpha) + \ell' R_\nu(\cos^2 \alpha) = \ell',$$

that is,

$$(*) \qquad R_\mu(\cos^2 \alpha) + R_\nu(\cos^2 \alpha) - 1 = 0.$$

Since R_μ a

constant ter

is an algebr

has four pre

$\cos^2 \alpha = \frac{1}{e}$.

any positive

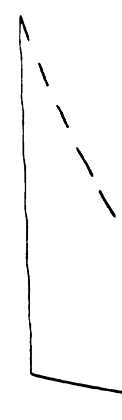
does not hav

satisfied.

Fig

geometric p

this space-



of curve

successo

constru

Since R_μ and R_ν are polynomial functions with no constant terms, (*) can only be satisfied if $1/\cos^2 \alpha$ is an algebraic integer. Thus to insure that no point has four pre-images one could, for example, choose $\cos^2 \alpha = \frac{1}{e}$. In fact, it suffices for $\cos^2 \alpha$ to be any positive fraction $< 1/\sqrt{2}$ which, in reduced form, does not have 1 as its numerator and (*) will not be satisfied.

Figure 13 represents Polya's version of the geometric picture of the first three approximations to this space-filling curve.

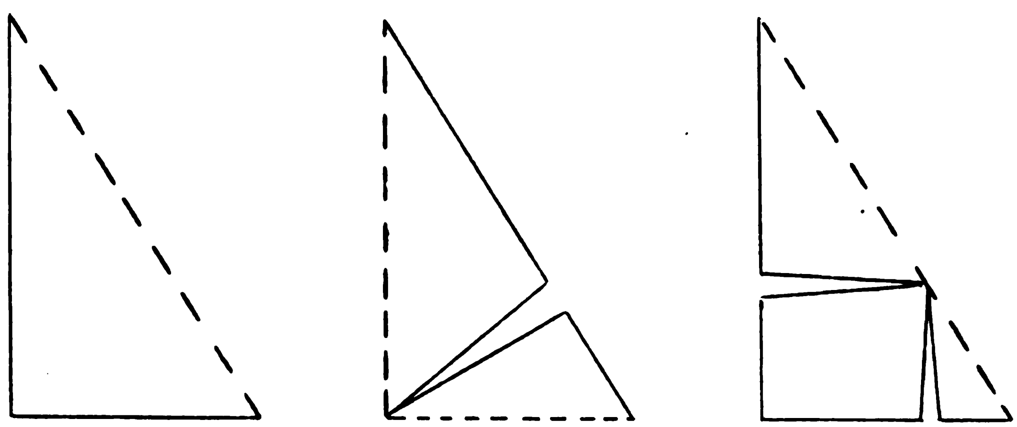


Figure 13

Some other related questions about the nature of curve also arise out of the work of Cantor and his successors. One such question is whether or not one can construct a one-to-one function from the unit interval onto the

where on

The answ

in 1913

of π c

perfect

function

subset

mentary

d into

P_c and

of the

the val

t in

determi

and d_c

with en

Similar

by P_1

with th

in p

onto the unit square given by parametric equations

$$x = \varphi(t) \quad \text{and} \quad y = \psi(t), \quad 0 \leq t \leq 1,$$

where one of them, say φ, is a continuous function of t. The answer to this question was given in the affirmative in 1913 by H. Hahn [20].

Hahn's construction involves assigning the values of φ on a sequence of mutually disjoint, nowhere dense, perfect subsets of the unit interval, and extending this function to the whole unit interval in a very natural way.

Let P be an arbitrary nowhere dense perfect subset of the unit interval. Let d denote the complementary intervals determined by P. Separate the intervals d into two nonempty sets of intervals d_o and d_1. Let P_o and P_1 be arbitrary nowhere dense perfect subsets of the intervals d_o and d_1, respectively. Assign the value $\varphi(t) = \frac{1}{4}$ for t in P_o and $\varphi(t) = \frac{3}{4}$ for t in P_1. Now separate the complementary intervals determined by P_o into two collections of intervals d_{oo} and d_{o1}, with the requirement that all those intervals with endpoints in the set P be among the intervals d_{o1}. Similarly separate the complementary intervals determined by P_1 into two collections of intervals d_{1o} and d_{11} with the requirement that those intervals with endpoints in P be among the intervals d_{1o}.

Hahn

hypothesis,

n i

as w

such

rule

(k $\leq$

(wit

alwa

d_{i_1}

by

d_{i_1}

per

set

bor

the

k

$\frac{j_1}{2}$

fol

two

Hahn thus arrives at the following induction hypothesis, which we quote in its entirety:

> Suppose the intervals $d_{i_1 i_2 \ldots i_n}$ with n indices 0 and 1 are already defined, as well as the sets $P_{i_1 i_2 \ldots i_{n-1}}$ with $n-1$ such indices, and this according to the following rules:
>
> 1. Each interval $d_{i_1 i_2 \ldots i_{k-1} i_k}$ $(k \leq n)$ lies in an interval $d_{i_1 i_2 \ldots i_{k-1}}$ (with the same first $k-1$ indices).
>
> 2. The sets $P_{i_1 i_2 \ldots i_k}$ $(k \leq n-1)$ always lie in the interior of the intervals $d_{i_1 i_2 \ldots i_k}$ and these intervals are separated by them into the intervals $d_{i_1 i_2 \ldots i_k 0}$ and $d_{i_1 i_2 \ldots i_k 1}$.
>
> 3. Each set $P_{i_1 i_2 \ldots i_k}$ $(k \leq n)$ is perfect, nowhere dense, and no two of these sets have a point in common.
>
> 4. If an interval $d_{i_1 i_2 \ldots i_k}$ $(k \leq n)$ borders on a point of a set $P_{j_1 j_2 \ldots j_h}$ $(h \leq k)$, then the indices $i_1 i_2 \ldots i_k$ coincide with first k places of one of the two binary expansions of $\frac{j_1}{2} + \frac{j_2}{2^2} + \ldots + \frac{j_h}{2^h} + \frac{1}{2^{j+1}}$. From this it easily follows, that there exist only the following two possibilities for an interval $d_{i_1 i_2 \ldots i_k}$

purpose

is disj

That is

infinit

Conside

which b

of poin

cluster

But sin

$P_{j_1 j_2 \ldots}$

hood of

of some

point c

$P_{j_1 j_2 \ldots}$

this se

of P_{j_1}

$(k \leq n)$: (α) both of its endpoints belong to the same set $P_{j_1 j_2 \ldots j_h}$ $(h \leq k)$; (β) one of its endpoints belongs to the set $P_{i_1 i_2 \ldots i_{k-1}}$, the other to a set $P_{j_1 j_2 \ldots j_h}$ $(h \leq k)$. We further assume:

5. For every $k \leq n$ there are only finitely many intervals $d_{i_1 i_2 \ldots i_k}$ of type (β).

Condition 5 deserves further explanation. Its purpose is to insure that each of the sets $P_{i_1 i_2 \ldots i_n}$ is disjoint from all the other sets $P_{j_1 j_2 \ldots j_h}$ $(h \leq n)$. That is, suppose there were, for some $k \leq n$, an infinite number of intervals $d_{i_1 i_2 \ldots i_k}$ of type (β). Consider the collection of endpoints of these intervals which belong to $P_{i_1 i_2 \ldots i_{k-1}}$. This is an infinite set of points of the unit interval, hence it must have a cluster point p (also an element of $P_{i_1 i_2 \ldots i_{k-1}}$). But since there are only a finite number of sets $P_{j_1 j_2 \ldots j_h}$ $(h \leq k-1)$, this implies that every neighborhood of p must contain an infinite number of points of some one of these sets. That is, p is a limit point of $P_{j_1 j_2 \ldots j_h}$ for some $h \leq k - 1$. But $P_{j_1 j_2 \ldots j_h}$ is closed, therefore p is an element of this set, contradicting the fact that the intersection of $P_{j_1 j_2 \ldots j_h}$ and $P_{i_1 i_2 \ldots i_{k-1}}$ is empty.

to defin

$d_{i_1 i_2 \cdots}$

is assig

Notice t

an inte

interval

then i.

places

structi

be exte

the ent

by assi

$p_{i_1 i_2 \cdots}$

where,

definit

indicat

Using the induction hypothesis, Hahn proceeds to define the sets $P_{i_1 i_2 \ldots i_n}$ and the intervals $d_{i_1 i_2 \ldots i_n i_{n+1}}$. To each t in the set $P_{i_1 i_2 \ldots i_n}$ is assigned the value

$$\varphi(t) = \frac{i_1}{2} + \frac{i_2}{2^2} + \ldots + \frac{i_n}{2^n} + \frac{1}{2^{n+1}}.$$

Notice that the assignment of the subscripted indices to an interval $d_{i_1 i_2 \ldots i_n i_{n+1}}$ is made so that if the interval contains points of the set $P_{j_1 j_2 \ldots j_h}$ $(h \leq n)$, then $i_1, i_2, \ldots, i_n, i_{n+1}$ coincide with the first $n + 1$ places of one of the two binary expansions of

$$\frac{j_1}{2} + \frac{j_2}{2^2} + \ldots + \frac{j_h}{2^h} + \frac{1}{2^{h+1}}.$$

That φ is continuous is clear from the construction, and it is not difficult to see that φ can be extended in a most natural way to a surjection from the entire unit interval to itself. This is accomplished by assigning to each t, not in one of the sets $P_{i_1 i_2 \ldots i_n}$ for any n, the value

$$\frac{j_1}{2} + \frac{j_2}{2^2} + \ldots + \frac{j_k}{2^k} + \ldots$$

where, for all $k > 0$, t is in the interval $d_{j_1 j_2 \ldots j_k}$.

Although Hahn does not specify a corresponding definition of $\psi(t)$, a few observations suffice to indicate what an appropriate choice might be. First, note tha

whose ca

This is

But supp

where th

infinit

the int

note that for each x, $0 \leq x \leq 1$, $\varphi^{-1}(x)$ is a set whose cardinality is the cardinality of the continuum. This is clear if x has a finite binary expansion. But suppose x is given by the expansion

$$\frac{i_1}{2} + \frac{i_2}{2^2} + \ldots + \frac{i_n}{2^n} + \ldots,$$

where the sequence of i_k's does not terminate in an infinite succession of 0's or of 1's.

Hahn has this to say about how one can choose the intervals $d_{i_1}, d_{i_1 i_2}, \ldots, d_{i_1 i_2 \ldots i_n}, \ldots$:

> We select two intervals d_0 and call them $\delta_0^{(0)}$ and $\delta_0^{(1)}$, and likewise two intervals d_1, which we denote by $\delta_1^{(0)}$ and $\delta_1^{(1)}$; we insist that in each of the two intervals $\delta_0^{(0)}$ and $\delta_0^{(1)}$ be contained points of P_0 and in each of the two intervals $\delta_1^{(0)}$ and $\delta_1^{(1)}$ be contained points of P_1. From among the subintervals into which $\delta_{i_1}^{(k_1)}$ $(k_1, i_1 = 0,1)$ is separated by P_{i_1}, we consider two intervals $d_{i_1 0}$ and two intervals $d_{i_1 1}$, which we denote by $\delta_{i_1 0}^{(k_1 1)}$, and $\delta_{i_1 1}^{(k_1 0)}$ and $\delta_{i_1 1}^{(k_1 1)}$, respectively; again we insist that in each of these intervals

betwee

and th

$(k_1 k_2$
$i_1 i_2$

detern

the in

elemen

k_1, k_2

the ca

the co

x fr

fore c

$P_{j_1 j_2}$
analog

for ex

early

is nec

$\delta_{i_1 i_2}^{(k_1 k_2)}$ be contained points of the set $P_{i_1 i_2}$.
Proceeding in this was we arrive at intervals
$\delta_{i_1 i_2 \dots i_n}^{(k_1 k_2 \dots k_n)}$, such that in each interval
$\delta_{i_1 i_2 \dots i_n}^{(k_1 k_2 \dots k_n)}$ lie the four intervals $\delta_{i_1 i_2 \dots i_n 0}^{(k_1 k_2 \dots k_n 0)}$,
$\delta_{i_1 i_2 \dots i_n 0}^{(k_1 k_2 \dots k_n 1)}$, $\delta_{i_1 i_2 \dots i_n 1}^{(k_1 k_2 \dots k_n 0)}$, and $\delta_{i_1 i_2 \dots i_n 1}^{(k_1 k_2 \dots k_n 1)}$.

In this way one can set up a correspondence between the sequences of 0's and 1's, $k_1, k_2, \dots, k_n, \dots$, and the sequences of intervals $\delta_{i_1}^{(k_1)}, \delta_{i_1 i_2}^{(k_1 k_2)}, \dots,$ $\delta_{i_1 i_2 \dots i_n}^{(k_1 k_2 \dots k_n)}, \dots$. Each such sequence of intervals determines a distinct point p which is common to all of the intervals of the sequence, and furthermore p is an element of $\varphi^{-1}(x)$. But since the cardinality of sequences $k_1, k_2, \dots, k_n, \dots,$ is the cardinality of the continuum, the cardinality of $\varphi^{-1}(x)$ is also the cardinality of the continuum. Thus there exists a one-to-one function ψ_x from $\varphi^{-1}(x)$ onto the unit interval. We can therefore define ψ to be ψ_x on $\varphi^{-1}(x)$. On the sets $P_{j_1 j_2 \dots j_k}$ one can define ψ to be an appropriate analog of the well-known Cantor ternary function. (See, for example, [69, p.48].)

Another question which was asked at least as early as 1887 by Jordan is whether or not a Jordan curve is necessarily a set of measure zero. (By a Jordan curve

unit i

rectif

was nc

arbitr

curren

set.

$\{s_\alpha\}\alpha$

and i

of M

where

of M

We no

first

indepe

curve

remin

Yoney

Osgoo

curve we mean a one-to-one and continuous image of the unit interval.) Jordan showed that the "area" of a rectifiable curve is always null, but until 1903 it was not known whether or not the "exterior area" of an arbitrary Jordan curve was necessarily zero.

By the exterior area of a set is meant what we currently refer to as the (Lebesgue) outer measure of a set. That is, if one considers an arbitrary covering, $\{s_\alpha | \alpha \in I\}$, of a plane point set, M, by open squares, and if a_α is the area of s_α, then the outer measure of M is

$$\inf \sum_{\alpha \in I} a_\alpha ,$$

where this infimum is taken over all possible coverings of M by open squares. (See, for example, [69, p.54].) We note in passing that it was apparently Peano who first introduced the notion of exterior area around 1887.

In 1903, W. Osgood [64] and H. Lebesgue [47], independently gave almost identical examples of a Jordan curve whose outer measure is positive. The construction reminds one of the familiar lakes of Wada described by Yoneyama [84] in 1917. We give here a description of Osgood's version of the construction.

diagon
is ext
think
square
water
dig ca
14, di
we num
(again
water
strong
lines
segmen

Begin with a square in the plane, one of whose diagonals (say, through the vertices labeled A and B) is extended infinitely in either direction. One can think of the two regions thus created exterior to the square as being bodies of water -- one of them fresh water and the other salt water. One then proceeds to dig canals through the square as indicated in Figure 14, dividing the square into nine equal squares, which we number for convenience in describing the construction (again, see Figure 14). Where the fresh water and salt water canals meet inside the square, erect thin (but strong) dikes. These dikes are indicated by the bold lines in Figure 14. It is the dikes which will form segments of the Jordan curve under construction.

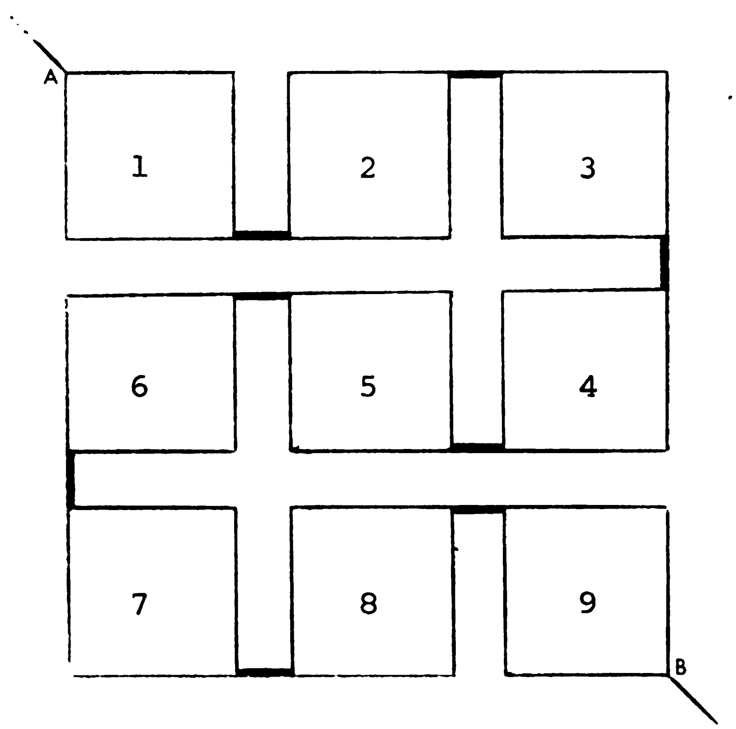


Figure 14

segme

$0 \leq t$

of sq

in th

l and

tinuo

the s

line

(2k-2

squar

is ma

k + l

of th

The r

the s

along

17 ec

an ar

along

cons

Now divide the unit interval into 17 equal segments. Starting at point A map the segment $0 \leq t < 1/17$ continuously along two consecutive sides of square 1. The segment $1/17 \leq t \leq 2/17$ is mapped in the obvious way onto the bold line connecting squares 1 and 2. The segment $2/17 < t < 3/17$ is mapped continuously along two consecutive sides of square 2, and the segment $3/17 \leq t \leq 4/17$ is mapped onto the bold line connecting squares 2 and 3. In general, the segment $(2k-2)/17 < t < (2k-1)/17$ is mapped along two sides of square k, while the segment $(2k-1)/17 \leq t \leq 2k/17$ is mapped onto the bold line connecting squares k and $k + 1$. This completes the first stage of the construction.

The next stage mimics the first stage, using each of the nine smaller squares in place of the larger square. The result is pictured in figure 15. This time each of the segments of the unit interval which were mapped along the sides of the numbered squares is divided into 17 equal segments, each of length $1/17^2$, and mapped in an analogous way onto the sides of the new squares and along the new bold lines introduced at this stage of the construction.

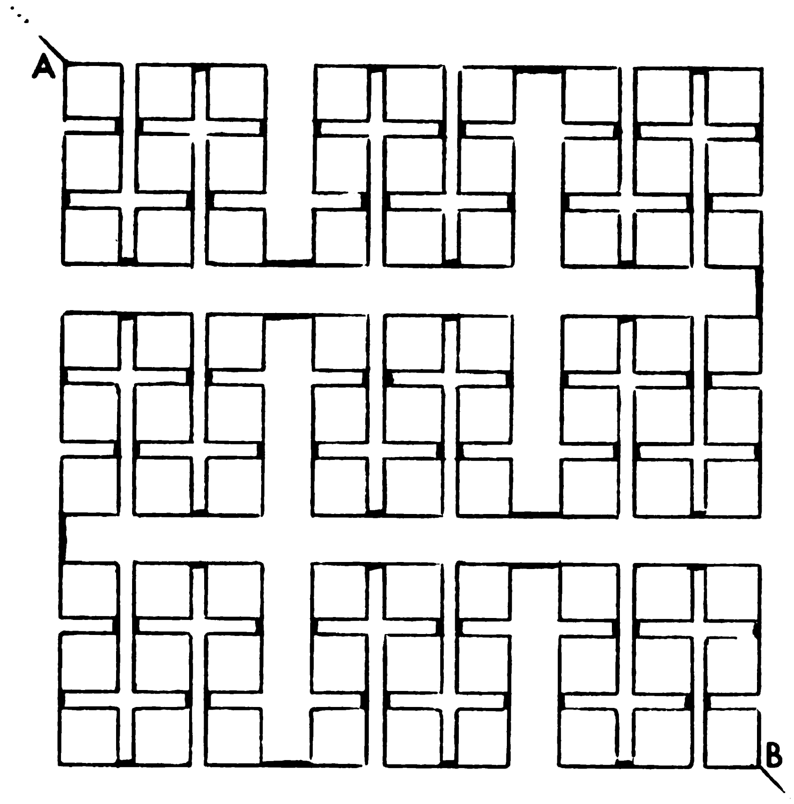


Figure 15

This construction is carried on ad infinitum. The desired Jordan curve C is then taken to be the union of all the bold lines of the construction together with the limit points of this union.

It follows immediately from the construction that C is indeed a Jordan curve and that every point of the square belongs to either the network of canals or to C. As yet, however, it is not clear that this curve has the desired property -- namely, that its outer measure is positive. This property is assured by selecting the widths of the canals at each stage of the construction as follows.

the kth

at that

in the

A

and in

Since

is gr

relat

which

prop

dens

a sp

But

cur

Rap

Osg

sub

Let ε be a positive number less than 1. At the kth stage, choose the width of the canals introduced at that stage to be $\frac{\varepsilon}{2\cdot 3^{2k-1}}$. The total area remaining in the square after these canals are dug is

$$A_k = 3^{2k}\cdot\left[\frac{1-\{2\cdot\frac{\varepsilon}{2\cdot 3}+2\cdot 3\cdot\frac{\varepsilon}{2\cdot 3^3}+\ldots+2\cdot 3^{k-1}\cdot\frac{\varepsilon}{2\cdot 3^{2k-1}}\}}{3^k}\right]^2$$

$$= [1-\{\tfrac{\varepsilon}{3}+\frac{\varepsilon}{3^2}+\ldots+\frac{\varepsilon}{3^k}\}]^2$$

and in the limit A_k tends to

$$\left[1-\sum_{k=1}^{\infty}\frac{\varepsilon}{3^k}\right]^2 = [1-\tfrac{\varepsilon}{2}]^2.$$

Since $1-\frac{\varepsilon}{2} > \frac{\varepsilon}{2}$, we have that the outer measure of C is greater than $\frac{\varepsilon^2}{4}$, which is positive. We note the close relationship between C and the generalized Cantor set, which can have arbitrary positive measure less than 1.

Even though C possesses the very unsettling property of having positive measure, it is still a nowhere dense subset of the square, and hence cannot be considered a space-filling curve in any reasonable sense of the word. But note the similarity with the construction of the Peano curve. As Schoenflies ([71, p.256]) observed, if one maps the "discarded" segments of the unit interval in Osgood's construction to a diagonal of each of the subsquares instead of along two sides of those squares,

the res
with m

curv
divi
bou
a q
ver
on
on
qu
de
ab
pr

the resulting figure is precisely that of the Peano curve with multiple points pulled apart (see Figure 16).

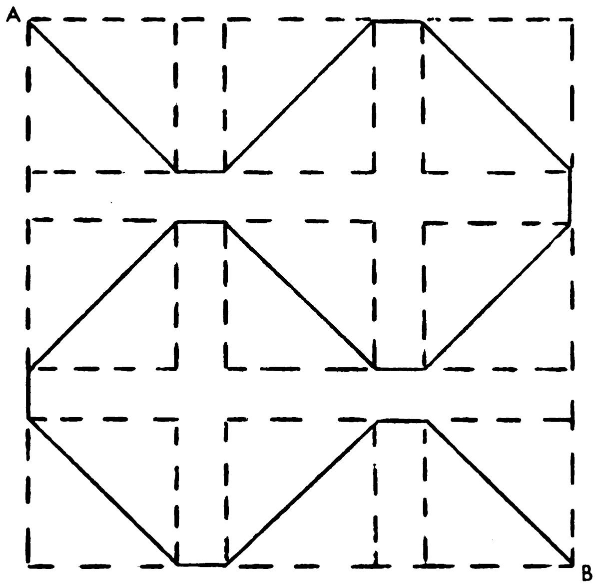


Figure 16

Osgood was also able to construct a simple closed curve having the same property. This is accomplished by dividing an annulus into an even number of equal sectors bounded by radii. In one of these sectors is constructed a quadrilateral (see Figure 17) such that no two of its vertices lie on the same radius and with one vertex, A, on one side of the sector and the opposite vertex, B, on the other side. The curve C is constructed in this quadrilateral in a manner analogous to the Jordan curve described above. The resulting curve is then reflected about the radius through the point A, and this reflection process is continued until the entire annulus is traversed.

Ar

positive r

Young [85

some gene

curves.

(

(

H

misleadir

basic ide

regions c

their co

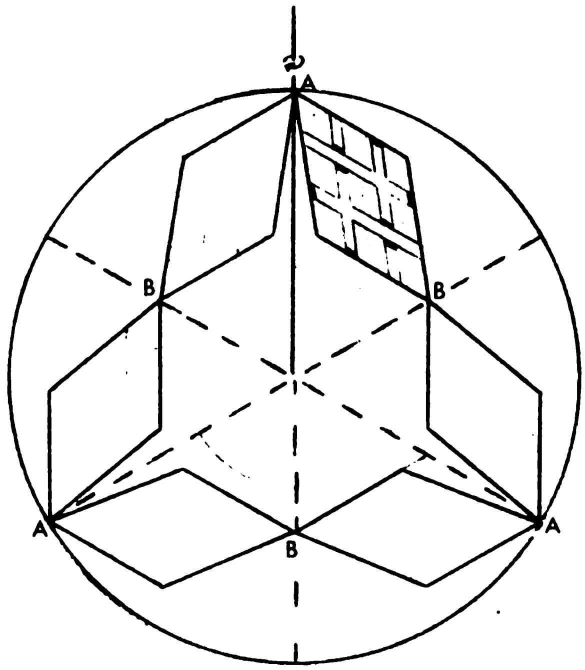


Figure 17

Another example of a simple closed curve having positive measure was given in 1905 by Grace Chisholm Young [85]. In this same paper Mrs. Young points out some general principles behind the construction of such curves. First she makes the following observations:

(1) A simple closed curve is the boundary of a simply connected region;

(2) If one joins two points of a Jordan curve by another Jordan curve, a simply connected region is enclosed.

Her attempt to give a general method is somewhat misleading so we shall only present her example. The basic idea is to construct two simply connected sub-regions of a bounded simply connected region such that their common boundary has positive measure.

in fac

square

three

X_1, a

a band

with

by the

one a

co

th

The procedure Mrs. Young uses is similar to (in fact, based on) that used by Osgood. Begin with a square region in the plane. Divide this region into three subsets (disjoint except for their boundaries) -- X_1, a cross with base on one side of the square, Y_1, a band of uniform width around the perimeter of the square (with the exception of that part of the perimeter occupied by the cross), and S_1, the union of four small squares, one at each corner of the cross (see Figure 18).

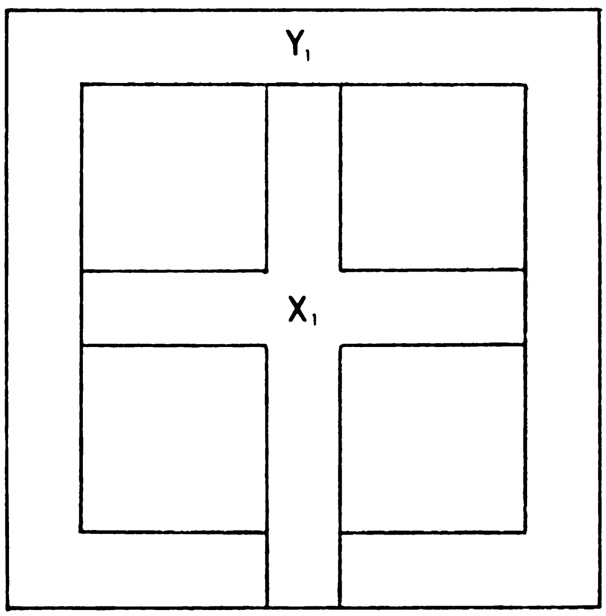


Figure 18

Figure 19 illustrates the second stage of the construction. In each of the squares which make up S_1, the construction of the first stage is repeated, the only

added
lie or
new c
will
bands
Int(Y
of th
This
that
the
that

added requirement being that the base of the cross must lie on the boundary of X $(= X_1)$. If we label the four new crosses, $X_2^{(1)}, X_2^{(2)}, X_2^{(3)}, X_2^{(4)}$, then $\mathrm{Int}(X_1 \cup \bigcup_{i=1}^{4} X_2^{(i)})$ will be denoted X_2. Similarly, if the four new uniform bands are labeled $Y_2^{(1)}, Y_2^{(2)}, Y_2^{(3)}, Y_2^{(4)}$, then $\mathrm{Int}(Y_1 \cup \bigcup_{i=1}^{4} Y_2^{(i)})$ will be denoted Y_2, and the union of the sixteen small square regions will be denoted S_2. This construction is carried out ad infinitum. Note that the widths of the uniform bands and the widths of the segments of the crosses can be chosen in such a way that the intersection $\bigcap_{i=1}^{\infty} S_i$ has positive area.

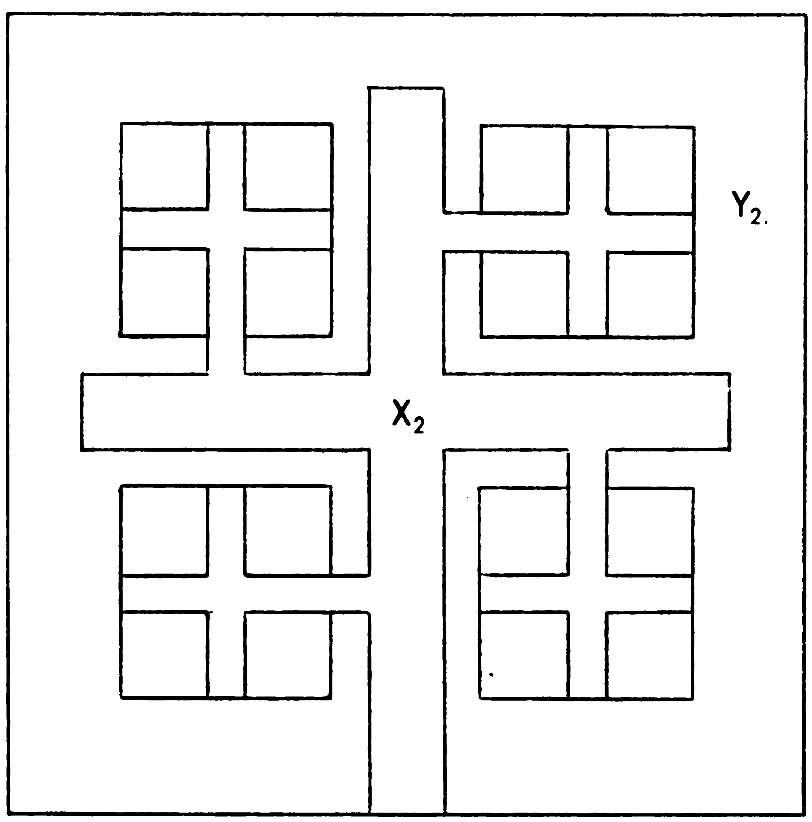


Figure 19

E

Lebesgue

in 1913 c

Sierpińsk

segment c

in Osgood

S

T, with

has leng

denoted I

unique i

on AC

triangle

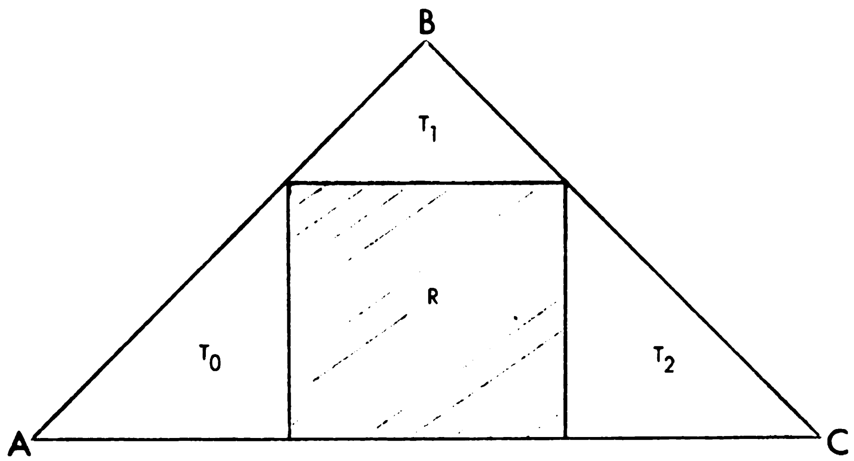
contain

A

Evidently unaware of the earlier papers of Osgood, Lebesgue and Young, Sierpiński [78] published an example in 1913 of a Jordan curve having positive measure. Sierpiński's curve has the additional property that any segment of it also has positive measure. (The "dikes" in Osgood's construction each have measure zero.)

Sierpiński starts with a right isoceles triangle, T, with vertices A, B, and C, such that the base AC has length less than 1. The area of the triangle is denoted by $\bar{T}$. It is easy to verify that there is a unique inscribed rectangle, R, with its shortest side on AC and having area, $\bar{R}$, equal to $\bar{T}^2$.

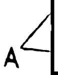
Removing R from T leaves the three smaller triangles, T_0 (containing the vertex A), T_1 (containing the vertex B), and T_2 (containing the



vertex C)

triangle

rectangle



T

particula

of the tr

$T_{i_1 i_2 \dots i}$

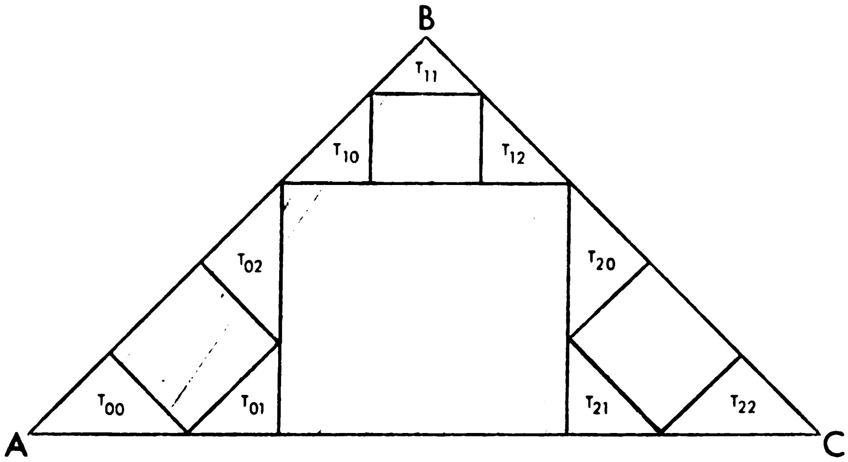
$T_{22\dots21'}$

consistin

of the ve

common ve

vertex C), each of which is similar to triangle T. In triangle T_i $(i = 0,1,2)$ is inscribed the unique rectangle R_i (as above) with area $\bar{R}_i$, equal to $\bar{T}_i^2$.



This process is repeated indefinitely. In particular, at the kth stage the configuration consists of the triangles $T_{00\ldots00}, T_{00\ldots01}, T_{00\ldots02}, T_{00\ldots01}, \ldots,$ $T_{i_1 i_2 \ldots i_{k-1}0}, T_{i_1 i_2 \ldots i_{k-1}1}, T_{i_1 i_2 \ldots i_{k-1}2}, \ldots, T_{22\ldots20},$ $T_{22\ldots21}, T_{22\ldots22}$. Each finite sequence, $i_1, i_2, \ldots, i_k$, consisting of 0's, 1's and 2's, corresponds to one of the vertices of the triangle $T_{i_1 i_2 \ldots i_k}$, namely the common vertex of all the triangles in the sequence

$$T_{i_1 i_2 \ldots i_k}, T_{i_1 i_2 \ldots i_k 0}, T_{i_1 i_2 \ldots i_k 00}, \ldots .$$

Simi

cons

the

sequ

Thus

from

$c(t)$

T_{i_1},

term

This

dist

unit

conv

one-

to t

corr

1,0,

i_1, i

of t

c i

Jord

Similarly, to each infinite sequence, $i_1, i_2, \ldots, i_k, \ldots,$ consisting of 0's, 1's and 2's, there corresponds the unique point common to all the triangles in the sequence

$$T_{i_1}, T_{i_1 i_2}, T_{i_1 i_2 i_3}, \ldots, T_{i_1 i_2 \ldots i_k}, \ldots .$$

Thus Sierpiński arrives at a very natural mapping φ from the unit interval to the triangle T by defining $\varphi(t)$ to be the unique point common to the triangles $T_{i_1}, T_{i_1 i_2}, T_{i_1 i_2 i_3}, \ldots, T_{i_1 i_2 \ldots i_k}, \ldots,$ where one of the ternary representations of t is

$$\frac{i_1}{3} + \frac{i_2}{3^2} + \frac{i_3}{3^3} + \ldots + \frac{i_k}{3^k} + \cdots .$$

This function is well-defined since, if there are two distinct ternary representations of a given point of the unit interval, the corresponding sequences of triangles converge to the same point of the triangle T. It is also one-to-one. That is, two sequences of triangles converge to the same point of T if and only if one of them corresponds to the sequence of indices $i_1, i_2, \ldots, i_k,$ $1, 0, 0, \ldots,$ while the other corresponds to the sequence $i_1, i_2, \ldots, i_k, 0, 2, 2, \ldots,$ for some $i_1, i_2, \ldots, i_k,$ elements of the set $\{0, 1, 2\}$. Furthermore, as is easily shown, φ is continuous, and therefore the image of φ is a Jordan curve.

I's a

that

and s

It fo

Also

small

side

then

Sin

Bec

tha

tha

It

If α represents any finite sequence of 0's, 1's and 2's, then it is apparent from the construction that

$$\bar{T}_{\alpha 0} = \bar{T}_{\alpha 2}$$

and since

$$\bar{T}_{\alpha 0} + \bar{T}_{\alpha 1} + \bar{T}_{\alpha 2} < \bar{T}_{\alpha}$$

it follows that

$$\bar{T}_{\alpha 0} < \frac{1}{2}\,\bar{T}_{\alpha} \quad \text{and} \quad \bar{T}_{\alpha 2} < \frac{1}{2}\,\bar{T}_{\alpha}.$$

Also by construction, the hypotenuse of $T_{\alpha 1}$ is the smaller side of the rectangle R_{α} whereas the longer side of R_{α} forms one of the legs of $T_{\alpha 0}$. It follows, then, that

$$\bar{T}_{\alpha 1} < \frac{1}{2}\,\bar{T}_{\alpha}.$$

By a simple induction, $\bar{T}_{i_1 i_2 \ldots i_n} < \frac{1}{2^n}\,\bar{T}$ and so,

$$\bar{R}_{i_1 i_2 \ldots i_n} = \bar{T}^2_{i_1 i_2 \ldots i_n} < \frac{1}{4^n}\,\bar{T}^2.$$

Since there are 3^n rectangles removed at the nth stage,

$$\bar{S} = \sum_{\alpha} \bar{R}_{\alpha} < \sum_{n=0}^{\infty} \left(\frac{3}{4}\right)^n \bar{T}^2 = 4\bar{T}^2.$$

Because the hypotenuse AC was chosen to have length less than 1, $\bar{T} < \frac{1}{4}$ and therefore, $\bar{S} < \bar{T}$. This proves that the Jordan curve determined by φ has positive measure. It follows from the fact that all of the triangles are similar an

there is l

segment be

measure.

Tr

chapter ha

each cont

at any of

example o

Koch in 19

from a st

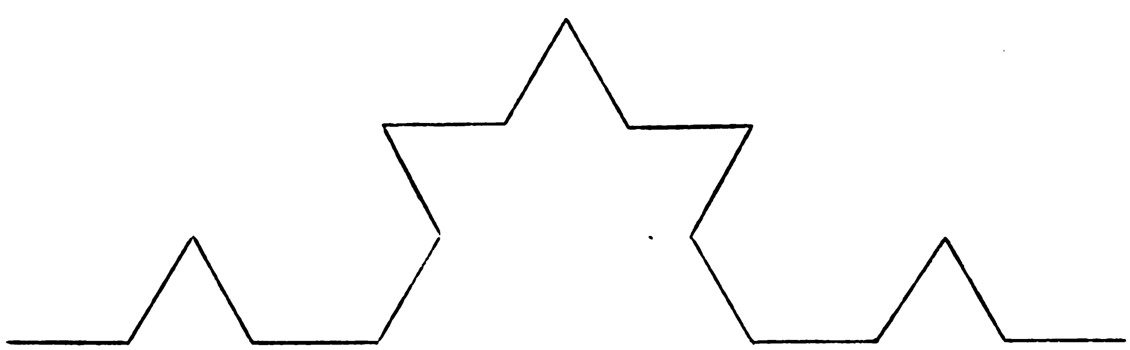
third of

with each

repeating

similar and that between any two points of the curve there is located a triangle T_α, for some α, that the segment between any two points of the curve has positive measure.

The examples of pathological curves given in this chapter have another property in common -- namely they each contain connected subsets, which have no tangent lines at any of their points. One well-known geometrical example of a tangentless curve was given by Helge von Koch in 1906 [43]. The von Koch curve is constructed from a straight line segment by "bumping out" the middle third of the segment to form a point and then, starting with each of the four resulting straight line segments, repeating this process indefinitely.



An

with an is

B of 120

having one

This separ

triangles,

process wi

indefinite

stage of t

the von Ko

As

type curve

using this

any obtuse

angle, and

manner si

such that

converges

This is,

constructi

The Peano

the const

A

of Jordan

sets whic

Another way to construct this curve is to start with an isoceles triangle ABC with an obtuse angle at B of 120°, and remove from it an equilateral triangle having one vertex at B and one side on the side AC. This separates triangle ABC into two smaller isoceles triangles, each similar to triangle ABC. Repeat the same process with each of the two new triangles and continue indefinitely. The limit set of the boundary at each stage of the construction will readily be recognized as the von Koch curve.

As Konrad Knopp [42] observed in 1917, Osgood type curves and Peano type curves can also be constructed using this same principle. For the Osgood curve select any obtuse triangle ABC, angle ABC being the obtuse angle, and remove from it obtuse sub-triangles in a manner similar to the construction of the von Koch curve such that the sum of the areas of the triangles removed converges to a value less than the area of triangle ABC. This is, of course, merely a modification of Sierpiński's construction which we have already discussed (see page 50). The Peano curve related to von Koch's curve is precisely the construction given by Polya (see page 29).

All of these examples illustrate the generality of Jordan's definition of curve. That is, the point sets which fall under this definition possess some rather unexpected properties -- properties not conforming to our common notion of what a curve should be. It was the study of examples such as the ones given in this chapter that set the stage for a precise point set characterization of "curve" in the sense of Jordan's definition. We shall consider this development in Chapter III.

to f

topo

and

par

Mat

Ent

([7

Sch

set

cur

of

gi

re

po

"i

wh

th

CHAPTER III

THE WORK OF SCHOENFLIES

One important work by Arthur Schoenflies seems to form the foundation for the study of curve from a topological viewpoint during at least the first decade and a half of the twentieth century. This is the two part paper which appeared in 1900 and 1908 in Deutsche Mathematiker Verein, Jahresbericht under the title, Die Entwicklung der Lehre von den Punktmannigfaltigkeiten ([71] and 775]). These two volumes were an attempt by Schoenflies to summarize the principal thrust of point set theory and develop a general theory of continuous curves as point sets. His results are done for the case of planar point sets, and then in Part II comments are given at the end of each chapter as to how some of these results might be extended to include higher dimensional point sets.

Much of the Entwicklung is devoted to discussing "invariants" -- characteristics of a general point set which are not altered by a continuous transformation of the set -- under the "wider group" (the continuous tran

"mar

itse

Scho

one

This

of t

The

diff

bega

were

clos

set

one

ther

curr

inv

Bro

thi

sim

Acc

a r

g e

transformations of the plane into itself) and under the "narrower group" (the homeomorphisms of the plane into itself). The need to find invariants was prompted, as Schoenflies explains, by Cantor's discovery of a one-to-one function from the unit interval to the unit square. This demonstrated the startling fact that the cardinality of the line is the same as the cardinality of the plane. The question thus arose: "What are the essential differences between the line and the plane?" And so began the search for invariants.

Among the invariants which Schoenflies discussed were connectedness (Zusammenhang), limit point (Grenzpunkt), closed and bounded set (abgeschlossene Menge), and perfect set. These are all invariant under the wider group. If one restricts one's attention to the narrower group, then the following properties are also invariant: closed curve, order of points on a curve, and region (the famous invariance of domain first proved in the general case by Brouwer [6] in 1911).

One important concept Schoenflies described in this work and then used to give a characterization of simple closed curves in the plane is that of accessibility. According to Schoenflies, a point m in the boundary of a region G is <u>accessible</u> from G if, for every point $g \in G$, there is a path joining m and g and contained entirely

m is sai

to G i

Fr(G),

contains

a point

accessib

Let T

half pla

point

regions

subset c

then m

accessil

every,

to both

entirely within G (with the exception of the point m). m is said to be <u>accessible from all sides</u> with respect to G if, for each arc α in G with endpoints in $Fr(G)$, and for each component G' of $G - \alpha$ which contains m in its boundary, m is accessible from G'.

The following example illustrates the fact that a point may be accessible from a region without being accessible from all sides with respect to that region. Let T be the "Warsaw circle" (situated in the right half plane as shown in Figure 20), and let m be the point $(0,-1)$. Then m is accessible from both of the regions of which T is the boundary, but if G' is the subset of the right half plane indicated in Figure 20, then m is contained in the boundary of G' but is not accessible from G'. If p is the point $(0,-2)$, however, then p is accessible from all sides with respect to both the unbounded region G and the bounded region I.

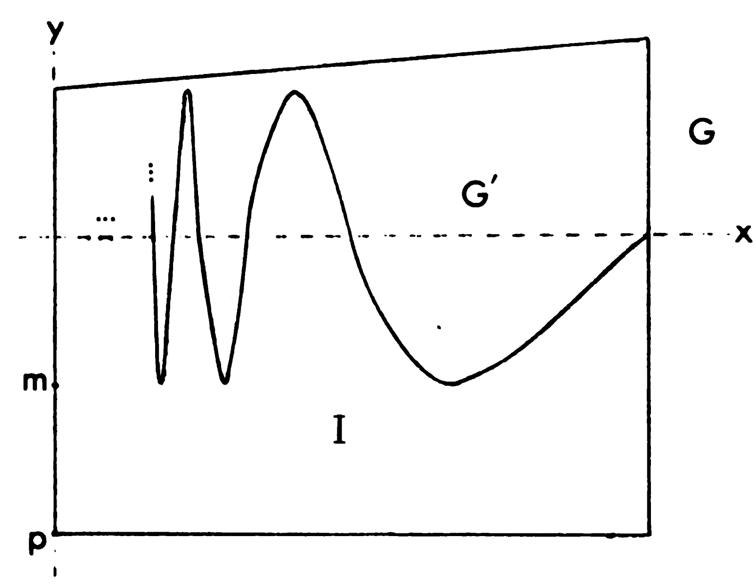


Figure 20

condi

plane

the p

point

to ea

find

(i)

the

sepa

sepa

lea

be

l_1

is

In

poi

by

mus

But

Thi

su

As Schoenflies proved, necessary and sufficient conditions for a closed and bounded subset M of the plane to be a simple closed curve are (i) that it separate the plane into exactly two regions, and (ii) that each point of M be accessible from all sides with respect to each of these regions. It was interesting to us to find that, in the presence of condition (ii), condition (i) can be weakened considerably. That is, because of the following result, we can replace (i) by (i′): M separates the plane into at least two regions.

Proposition 3.1. Condition (ii) implies that M separates the plane into at most two regions.

Proof: Suppose M separates the plane into at least three regions, R_1, R_2, and R_3. Let q and q' be two points of M. By condition (ii), there is a path ℓ_1 from q to q' contained in $R_1 \cup \{q,q'\}$ and there is a path ℓ_2 from q to q' contained in $R_2 \cup \{q,q'\}$. In the bounded region determined by $\ell_1 \cup \ell_2$ there is a point m_1 of M and in the unbounded region determined by $\ell_1 \cup \ell_2$ there is a point m_2 of M. Thus, there must be a path ℓ_3 in $R_3 \cup \{m_1,m_2\}$ from m_1 to m_2. But ℓ_3 must intersect $\ell_1 \cup \ell_2$ in at least one point. This, however, is a contradiction, since $\ell_1 \cup \ell_2$ is a subset of $R_1 \cup R_2 \cup (M - \{m_1,m_2\})$. □

One immediately asks if conditions (i) and (ii) also characterize spheres of higher dimension than 1. The answer is no, as the example of a torus in 3-space clearly demonstrates. What further conditions, then, must one impose on a set to insure that it be a sphere? With the torus in mind, a logical requirement might be that the set have the same homology as a sphere.

Perhaps tacitly assuming such an added hypothesis, L.E.J. Brouwer attempted in 1911 to produce a counter-example (see [8]). A slightly modified version of his example, which we shall call the "Brouwer sphere", B, can be expressed as the union of the following sets of points in cylindrical coordinates:

$$\{(\rho,\theta,z): 0 < \rho < 1,\ z = 2 + \cos\theta + (1 - \sqrt{\cos^2\theta})\sin(\pi/\rho)\};$$
$$\{(1,\theta,z): 0 \leq z \leq 2 + \cos\theta\};$$
$$\{(0,\theta,z): 1 \leq z \leq 3\};$$
$$\{(\rho,\theta,0): 0 \leq \rho \leq 1\}.$$

Various cross-sections of B through the z-axis are given in Figure 21.

Brouwer claims that at the point H with coordinates $(0,0,2)$ there is a singularity in the sense that there exists a sequence of points of B converging to H which cannot be joined by a simple arc lying entirely in B. This claim is certainly not true. To

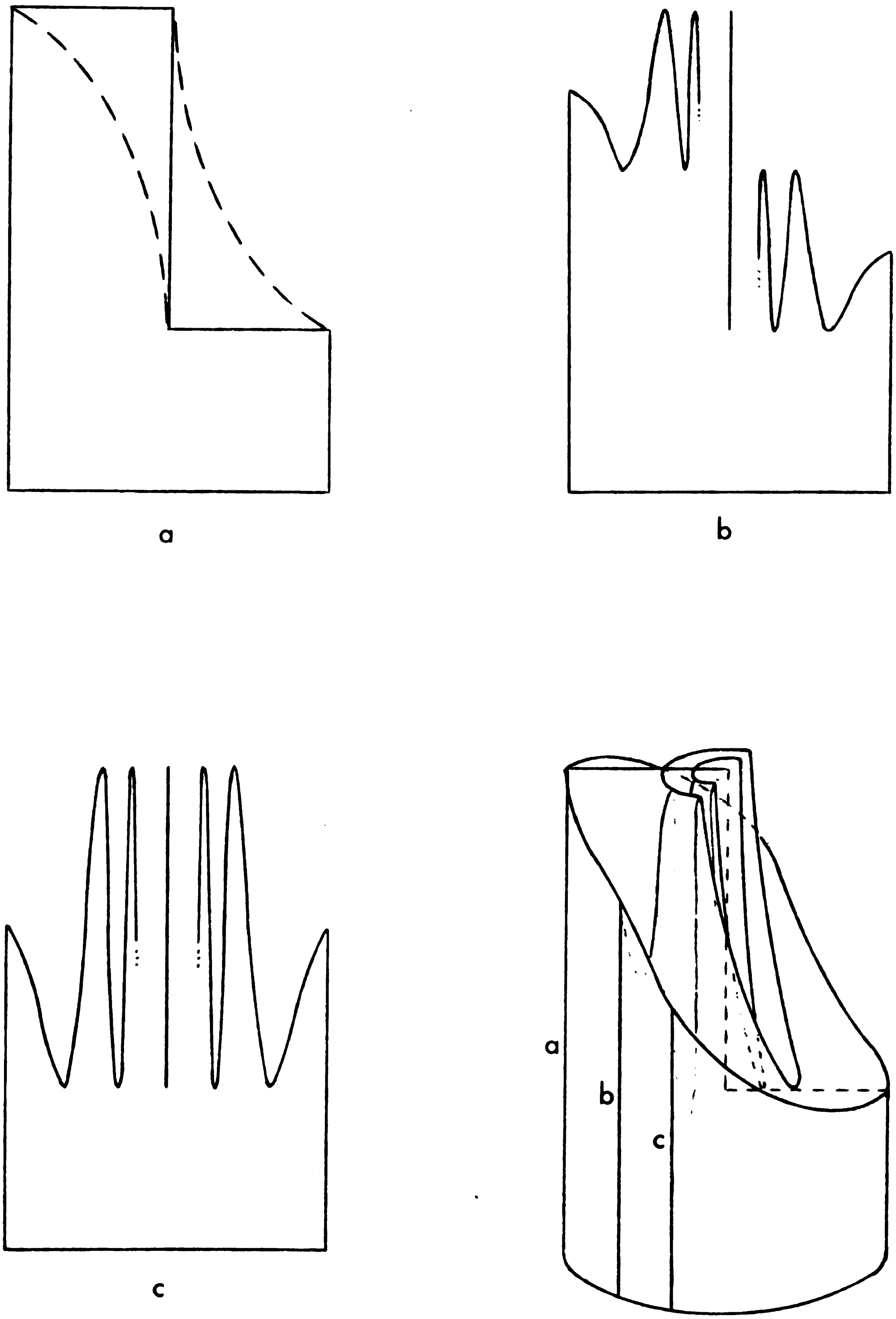


Figure 21

see wh

levels

conver

$2 \pm \frac{1}{3}$,

be at

These

entir

of th

can p

simpl

conta

the f

seque

taini

level

level

const

of th

furth

not a

and p

will

throu

see why, first separate B into a countable number of levels by planes parallel to the base plane $z = 0$ and converging to the plane $z = 2$ (for example, $z = 2 \pm \frac{1}{2}$, $2 \pm \frac{1}{3}$, $2 \pm \frac{1}{4}, \ldots$). Within any given level there can be at most a finite number of points of the sequence. These points can be joined in B by a simple arc lying entirely in the given level and not intersecting any of the previously constructed arcs. Furthermore, one can pass from one level to any other level along a simple arc. Now, to construct the desired simple arc containing all the points of the sequence, start with the furthest level above H containing points of the sequence and construct a simple arc in that level containing all the points of the sequence within that level. Then pass along a simple arc to the furthest level below H containing points of the sequence and construct an arc there which contains all the points of the sequence within that level. Proceed next to the furthest level above H containing points of the sequence not already used in the construction. Alternating above and below H in this way, the arc under construction will eventually converge to the point H and will pass through all the points of the given sequence.

In

standard 2-

to exhibit

onto a plan

in two stag

a concave h

Equations f

$1 \leq z \leq 3,$

$0 < \rho \leq 1,$

$x = -$

and for

$x = \rho$

The only p

for which

Le

with coord

be a seque

ρ_i conver

$\varepsilon > 0$ be

such that

In fact, the Brouwer sphere is homeomorphic to the standard 2-sphere. To prove this it is clearly sufficient to exhibit a homeomorphism from the top of the cylinder onto a plane disc. Such a homeomorphism can be accomplished in two stages. First inject the top of the cylinder onto a concave hexagonal planar region. (See Figure 22.) Equations for this map are as follows: for $\rho = 0$, $1 \leq z \leq 3$, let $x = 0$, and $y = z - 2$; for $0 \leq \theta \leq \pi$, $0 < \rho \leq 1$, let

$$x = -\rho, \text{ and } y = (1 - \sqrt{\cos^2 \theta})\sin(\pi/\rho) + (1-\rho)\cos\theta;$$

and for $\pi < \theta < 2\pi$, $0 < \rho \leq 1$, let

$$x = \rho, \text{ and } y = (1 - \sqrt{\cos^2 \theta})\sin(\pi/\rho) + (1+\rho)\cos\theta.$$

The only points where continuity is in question are those for which $\rho = 0$.

Let p be a point of the Brouwer sphere B with coordinates $(\rho, \theta, z) = (0, 0, z_0)$, and let

$$\{p_i\}_{i=1}^{\infty} = \{(\rho_i, \theta_i, z_i)\}_{i=1}^{\infty}$$

be a sequence of points of B converging to p. Then ρ_i converges to 0 and z_i converges to z_0. Let $\varepsilon > 0$ be given. Then there is a positive integer N such that for all $i \geq N$, $\rho_i < \varepsilon/3$ and $|z_i - z_0| < \varepsilon/3$.

$\theta = 2''$

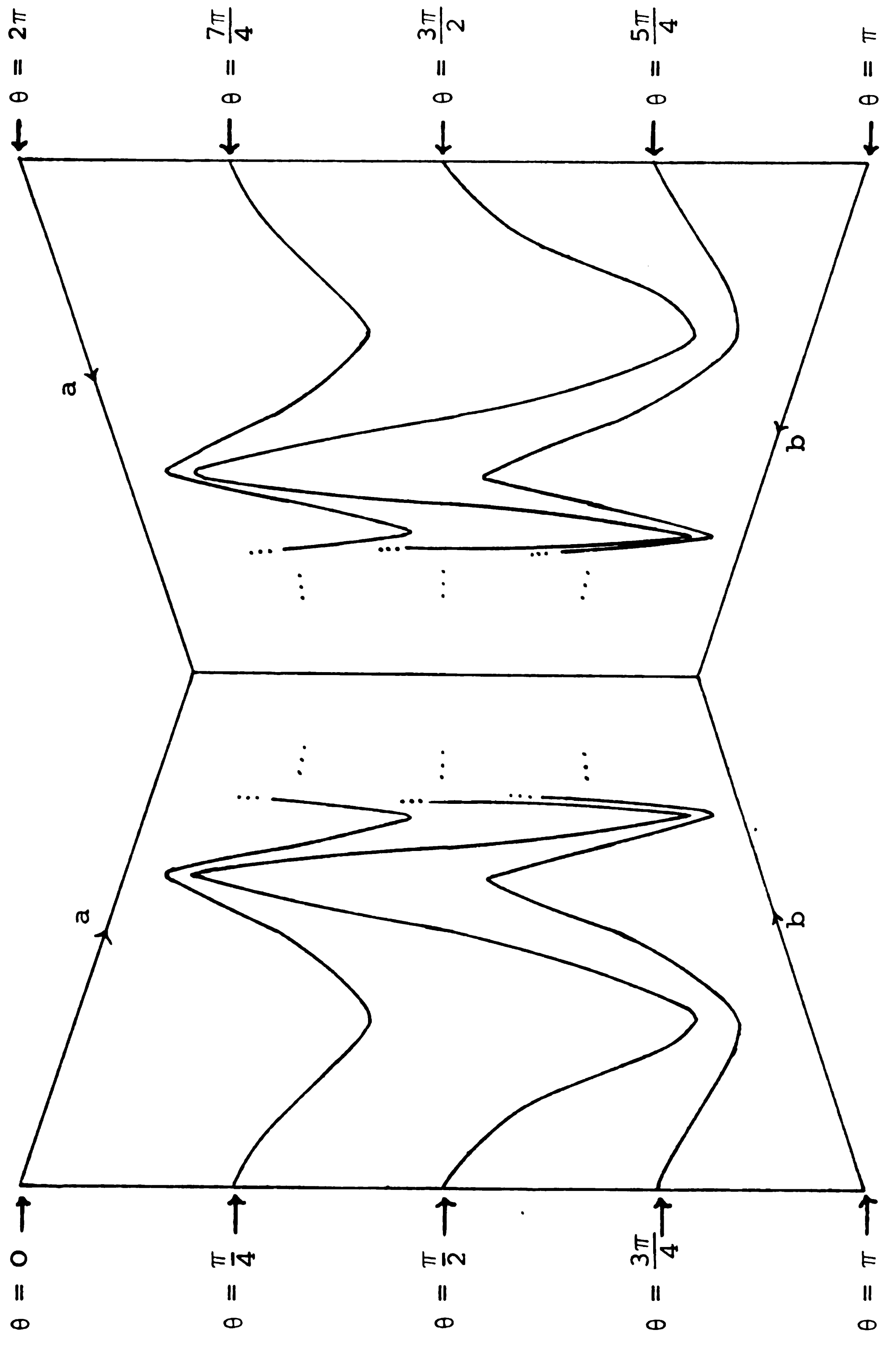


Figure 22

<u>Case 1</u>: $\rho_i = 0$. Then

$$d(f(p_i), f(p)) = |(z_i-2)-(z_o-2)| = |z_i-z_o| < \epsilon/3.$$

<u>Case 2</u>: $\rho_i > 0$ and $0 \leq \theta_i \leq \pi$. Then

$$d(f(p_i), f(p)) = \sqrt{\rho_i^2+[(1-\sqrt{\cos^2\theta_i})\sin(\pi/\rho_i)+(1-\rho_i)\cos\theta_i-(z_o-2)]^2}$$

$$= \sqrt{\rho_i^2+[(z_i-2-\rho_i\cos\theta_i)-(z_o-2)]^2}$$

$$= \sqrt{\rho_i^2+[(z_i-z_o-\rho_i\cos\theta_i]^2}$$

$$\leq \rho_i + |z_i-z_o| + |\rho_i \cos \theta_i|$$

$$< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

<u>Case 3</u>: $\rho_i > 0$ and $\pi < \theta_i < 2\pi$. Then

$$d(f(p_i), f(p)) = \sqrt{\rho_i^2+[(1-\sqrt{\cos^2\theta_i})\sin(\pi/\rho_i)+(1+\rho_i)\cos\theta_i-(z_o-2)]^2}$$

$$= \sqrt{\rho_i^2+[(z_i-2+\rho_i\cos\theta_i)-(z_o-2)]^2}$$

$$= \sqrt{\rho_i^2+[(z_i-z_o)+\rho_i\cos\theta_i]^2}$$

$$\leq \rho_i + |z_i-z_o| + |\rho_i \cos \theta_i|$$

$$< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

This establishes the continuity of the function at p.

The definition of the homeomorphism is completed by identifying the sides of the hexagon labelled "a" and those labelled "b" (see Figure 22) in the obvious way, to produce a disk.

Though the Brouwer sphere fails to be the counterexample Brouwer intended, the following example (due to P. Doyle) does have the same homology as the sphere and yet is not homeomorphic to the 2-sphere since it is not connected im kleinen. Consider the surface K of a cube with the Warsaw circle T embedded in one of its faces. All of the points of T are accessible from one of its complementary regions on K, but some fail to be accessible from the other. Delete the domain with respect to which accessibliity fails, and let K_o be the resulting surface. Now take another copy of K_o and identify the points of T on one copy of K_o with the corresponding points of T on the other. This gives us the desired counterexample. (See Figure 23 for a somewhat simplified version.)

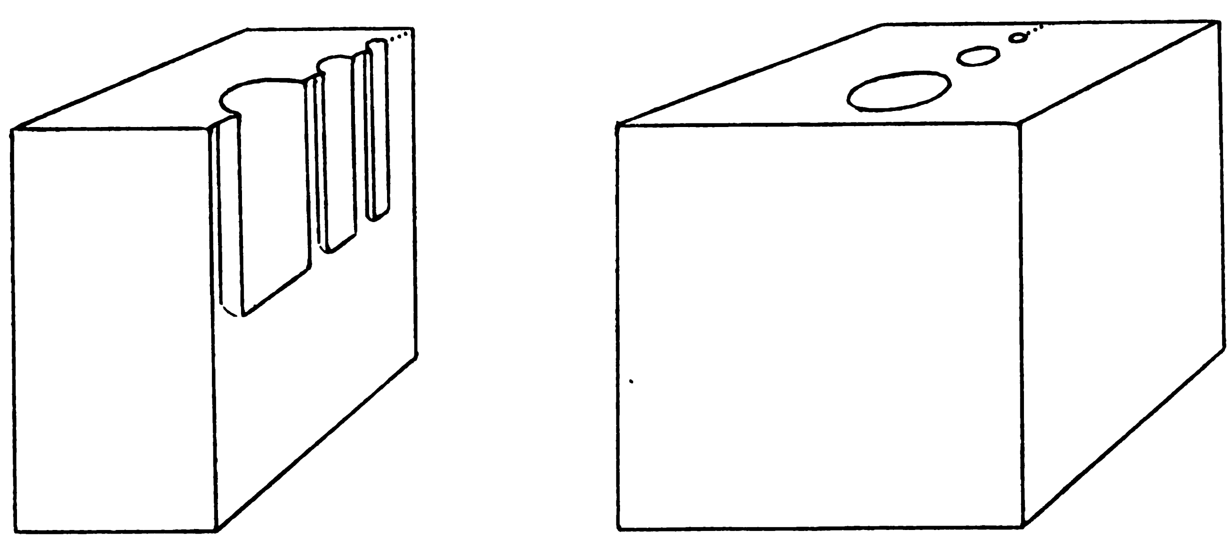


Figure 23

Perhaps the most important contribution contained in the Entwicklung, from the point of view of the present work, is Schoenflies' characterization of the most general point set in the plane which is a continuous image of the closed unit interval, that is, a continuous curve. Earlier we discussed the fact that Schoenflies was able to characterize the simple closed curve in the plane by means of the accessibility of its points. By adding one additional hypothesis to that of accessibility, Schoenflies arrived at the following result.

Theorem 3.2. A closed and bounded, connected, planar point set M is a continuous image of the unit interval iff the following conditions hold:

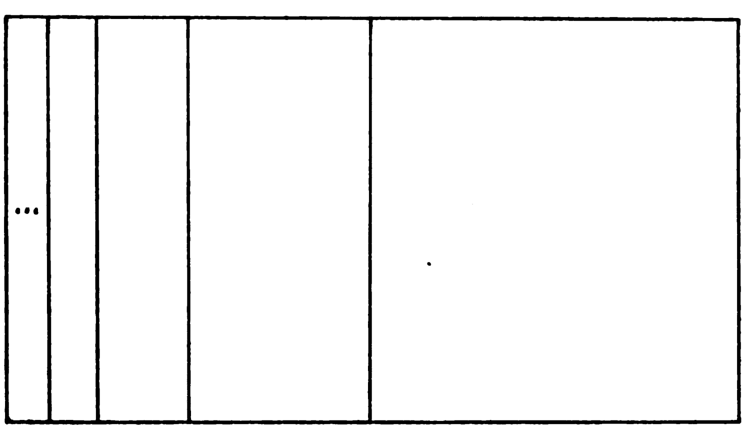
(i) If G is any component of the complement of M, and p is any limit point of G contained in M, then p is accessible from all sides with respect to G; and

(ii) If ϵ is an arbitrary positive number, the complement of M has at most a finite number of components whose diameters exceed ϵ.

Schoenflies divides the proof of Theorem 3.2 into four cases (see [75, p.200]). For historical interest we indicate here an outline of its development. (The entire proof is quite lengthy, taking up some 30 pages of Schoenflies' paper.)

The first step is to prove the theorem for the simplest case, the set M being the boundary of only one region and containing no surface-filling subsets. In the second case, M is assumed to contain one surface-filling piece. The next stage is to prove the theorem for the case where M separates the plane into an arbitrary (but finite) number of components and contains an arbitrary (but finite) number of distinct surface-filling pieces. Finally, the proof is demonstrated for the case where the complement of M contains an infinite number of components and/or there is an infinite number of distinct surface-filling pieces contained in M.

That condition (i) alone is not sufficient is shown by the following set:



The example of Figure 23 suffices to show that even with the addition of condition (ii) this characterization is not strong enough to cover sets of higher dimension than 2.

Accessibility is another invariant under the wider group in the following sense:

Theorem 3.3. (Schoenflies). If the boundary Γ of a planar region G is the continuous image of a circle or of a circular arc, then every point of Γ is accessible from all sides with respect to the region G.

Proof: See [75, p.189].

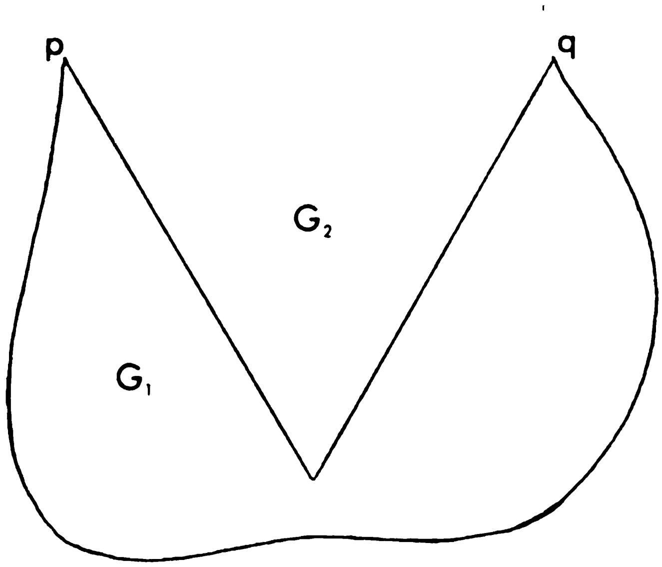
The concepts Schoenflies introduced to establish Theorem 3.3 -- namely, simple sequence of points and path-distance -- are of interest to us because of the influence they had on the development of uniform connectedness im kleinen (see Chapter IV).

To begin with, Schoenflies gave the following definition. Let $\{t_\nu\}_{\nu=1}^{\infty}$ be a sequence of distinct points of Γ having the unique limit point t_ω, such that with increasing ν, the distance between t_ν and t_ω decreases. Assume that the points t_ν (though not necessarily t_ω) are all accessible with respect to a region G. Thus each t_ν can be joined to a point m

f G by a path ℓ_ν, and, in fact, this can be done n such a way that no two of the paths have any other point than m in common. (The point m is assumed to be distinct from t_ω.) With m as center, construct a circle which contains no t_ν, either in its interior or on its boundary. This circle intersects the path ℓ_ν in the point k_ν. Using the order of the points on the circle, a subsequence of the k_ν's can be selected, say $\{k_{\nu_i}\}$, such that $\nu_i < \nu_{i+1}$ for all i, and taking a fixed orientation from the point k_o on the circle, k_{ν_i} precedes $k_{\nu_{i+1}}$ for all i. Corresponding to the sequence $\{k_{\nu_i}\}$, is the subsequence $\{t_{\nu_i}\}$ of the sequence $\{t_\nu\}$. $\{t_{\nu_i}\}$ is what Schoenflies called a simple sequence.

One immediate consequence of this definition is that every convergent sequence has a simple subsequence. Schoenflies used this fact implicitly in the proof of the theorem.

The second concept, path-distance, is important, not only for its role in the proof of Theorem 3.3, but also for the part it played in some of the work of Brouwer hinting at the concept of uniform connectedness im kleinen. Given two points p and q on the boundary of a region G, the path-distance (Wegdistanz) between p and q with respect to G is the infimum of the lengths f all paths in G connecting p and q. Thus, in he illustration below, the path-distance between p nd q is 2 with respect to G_1, but the path-istance between them is 1 with respect to G_2.



The accessible points of the boundary Γ of a simply connected region G possess the following properties:

(i) The points of Γ which are accessible from G are dense in Γ. (See [81, p.96].)

(ii) If $\{t_\nu\}$ is a simple sequence of points of Γ such that the path-distance η_ν (with respect to G) between t_ν and $t_{\nu+1}$ converges to zero with increasing ν, then the limit point t_ω is accessible with respect to G. (See [75, p.177].)

A direct consequence of (ii) is the following property.

(iii) Let $\{t_\nu\}$ be a simple sequence of points of Γ with t_ω as its limit point. Let m be a point of G such that for each ν there is a simple path ℓ_ν from m to t_ν. Then for any ν, the path $\ell_\nu \cup \ell_{\nu+1}$ separates G into two regions, I_ν and E_ν. Let I_ν be the one which contains neither $\ell_{\nu-1}$ nor $\ell_{\nu+2}$. Let Γ_ν be $\Gamma \cap \mathrm{Fr}(I_\nu)$. Then t_ω is accessible from G whenever the diameters of the sets Γ_ν converge to zero with increasing ν. (See [75, p.178].)

These three properties are the machinery Schoenflies used to prove Theorem 3.3.

The work of Schoenflies discussed in this chapter was, in a sense, a pioneering effort. In the next chapter we will present some of the results leading up to the Hahn-Mazurkiewicz Theorem -- results which, to a large degree, were built on the foundation laid by Schoenflies. We will also discuss briefly a definition of curve, due to Cantor, which appeals more to geometric intuition than does the definition given by Jordan.

Cantor's definition succeeds in excluding space-filling curves, but it still does not eliminate all pathology.

CHAPTER IV

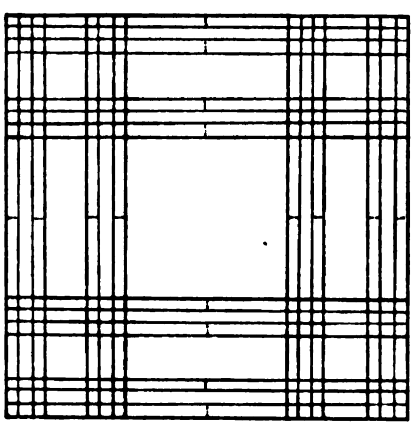
PRELUDE TO THE HAHN-MAZURKIEWICZ THEOREM

Between 1909 and 1913, Ludovic Zoretti published a series of 8 papers ([88],[89],[90],[91],[92],[93],[94], and [95]) in an attempt to clarify, and make precise, the concept of "line". Zoretti's express intention was to characterize those point sets which best generalize the notion of a simple arc. The discussion which follows is based primarily on two of these papers, [88] and [94], since the remainder of them contain only isolated results which were incorporated into [94].

In the early 1900's, there were two commonly used definitions of line, or curve -- one due to Cantor and the other to Jordan. According to Cantor, a continuum is any perfect, well-chained point set. A set F is well-chained if for every two points a and b of F and for every positive real number ϵ, there is a set of points $\{s_0 = a, s_1, \ldots, s_{k-1}, s_k = b\}$ contained in F with the distance between s_i and s_{i+1} less than ϵ for $i = 0,1,\ldots,k-1$. A Cantor line is defined to be a linear continuum, that is, a continuum having no interior points. On the other hand, a line, or curve, according to Jordan, is any set which is the continuous image of a closed line segment.

In [88], Zoretti asked the question: "How are these two definitions related?" It was already well-known that they are not identical notions, and that, in a sense, they are independent. Consider, for example, the familiar topologist's sine wave (the curve $\sin(\pi/x)$, for $0 < x \leq 1$, with the addition of its limit points on the y-axis). This fails to be a Jordan line since it is not a continuous image of a line segment, a fact with which Zoretti was familiar. But it is a Cantor line as is easily seen. An example of a Jordan line which is not a Cantor line is the unit square, since it has a non-empty interior.

Zoretti also suggested the following example of a Cantor line:



To construct the example, take two perfect, nowhere dense subsets of the segments $0 \leq x \leq 1$ of the x-axis and $0 \leq y \leq 1$ of the y-axis. At each point of these nowhere dense sets, erect a perpendicular (of length one) to each axis. The fact that the sets were chosen to be perfect insures us that the result will be a Cantor line.

Zoretti conjectured that it is highly unlikely that this example is also a Jordan line. This conjecture is easily shown to be true if one has at his disposal the work of Hahn and Mazurkiewicz, which we will examine later, but Zoretti's own work falls short of being able to prove it.

The next question Zoretti asked was, "What are the minimum conditions that one needs to add to each of these definitions of line in order that they define the same concept?" He then restricted his attention to finding the equivalent in Cantorean terms of the notion of a simple Jordan line, that is, one without multiple points.

[88] would almost assuredly have fallen into obscurity except for the fact that Zoretti here introduced, for the first time in print, the notion of an

irreducible continuum. A continuum C is said to be irreducible between two of its points, a and b, if there is no proper subcontinuum of C containing both a and b. We say simply that C is an irreducible continuum if there are points a and b in C such that C is irreducible between a and b. It is within the class of irreducible continua that Zoretti saw some hope of finding a natural definition of line -- one which fits our intuition of what a line should be. This search led him to define the notion of a simple irreducible continuum. An irreducible continuum C is simple if there do not exist two distinct pairs of points, (a,b) and (c,d), such that C is irreducible between a and b and at the same time between c and d. Otherwise C is said to be non-simple. For example, the closed unit interval along the x-axis in E^2 is a simple irreducible continuum between (0,0) and (1,0). The topologist's sine wave mentioned above is irreducible between the point (1,0) and any point on the y-axis between $y = -1$ and $y = 1$, and is, therefore, a non-simple irreducible continuum.

In his original paper ([88]) Zoretti had tried to get at this same concept in a somewhat different way, by defining what he meant for an irreducible continuum to be absolutely closed. An irreducible continuum C ($\subset E^2$) is said to be absolutely (or completely) closed, if, for every point m of C, given any disk D (in E^2) with center at m, there is a disk D' with center at m such that whenever p and q are in $C \cap D'$ and the irreducible subcontinuum $\overline{pq}$ of C between p and q contains m, then $\overline{pq}$ is contained in D. In [88], Zoretti proved that given two points p and q of an irreducible continuum, there exists a unique irreducible subcontinuum between p and q, which he called the "arc" $\overline{pq}$. This definition of arc has the disadvantage in the general case of not allowing the establishment of a linear ordering of its points. However, for simple irreducible continua, Zoretti's definition corresponds to the usual definition of a simple arc. Zoretti also proved ([94, p.251]) that an irreducible continuum is absolutely closed iff it is simple.

Observe that the form of the definition of absolute closure is remarkably close to that of connectedness im kleinen. This leads us to ask, "Are the two ideas identical in the setting of irreducible continua?"

We first note that, in general, a set may be well-chained and yet not be connected. For example, consider the set $S = \{(x, 1/x^2) : x \neq 0\}$. S is an example of a continuum in the sense of Cantor which is not a continuum in the sense of Jordan. (Jordan replaces the condition of being well-chained by that of being "d'un seul tenant," which, for closed sets, is equivalent to connectedness.) Note that S is unbounded. The set of all rational numbers in the interval $[0,1]$ is an example of a bounded set which is well-chained and yet is totally disconnected. The rationals, however, fail to be a continuum even in the sense of Cantor since they do not constitute a closed set.

The following well-known theorem, first proved by Jordan ([36, p.25]), though in a less general setting than is given here, shows that, under certain restrictions, being well-chained is equivalent to being connected.

Theorem 4.1. A compact subset C of a metric space is well-chained iff it is connected.

In the proof we use the following theorem:

Theorem 4.2. If p and q are two points of a connected space C, and $\{O_\alpha\}$ is a collection of open (in C) sets covering C, then there is a finite subcollection $\{O_i\}_{i=1}^{k}$ of elements of $\{O_\alpha\}$ such that $p \in O_1$, $q \in O_k$, and $O_i \cap O_j$ is nonempty iff $|i - j| \leq 1$.

Proof: See [29, p.108].

Proof of Theorem 4.1. $(\Rightarrow)$. If C is not connected, then C is the union of two disjoint closed subsets C_1 and C_2, of C. Each is compact and so they are a finite distance $\varepsilon > 0$ apart. Let $p \in C_1$, and $q \in C_2$. Then there does not exist an $\varepsilon/2$-chain between p and q. Therefore, C is not well-chained.

($\Leftarrow$). It is an immediate consequence of Theorem 4.2 that every connected subset of a metric space is well-chained. $\square$

In answer to the question of whether or not absolute closure and connectedness im kleinen are the same we prove the following proposition.

Proposition 4.3. Let C be an irreducible continuum. Then C is absolutely closed iff it is connected im kleinen.

Proof: ($\Rightarrow$). Suppose C is absolutely closed. Let m be a point of C and let D be a disk with center at m. Let D' be the disk corresponding to D in the definition of absolute closure, and let p be any point of $C \cap D'$. Then the arc $\overline{pm}$ is contained in D. Thus every point of D' lies in a connected subset of C containing m and contained in D. That is, C is connected im kleinen at m.

($\Leftarrow$). This direction can be more easily proved with the help of the following theorem.

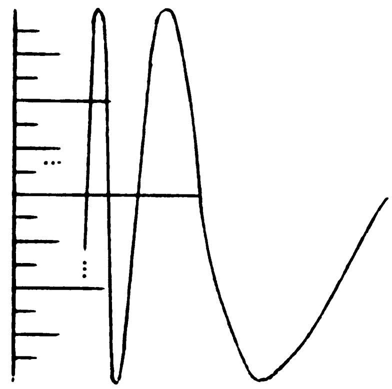
<u>Theorem 4.4</u>. (Janiszewski). Every continuum contains an irreducible subcontinuum between any two of its points.

<u>Proof</u>: See [30, p.606], or [48, p.296].

Suppose C is connected im kleinen at m. Then let $\epsilon > 0$ be given and choose $\delta > 0$ such that for p in $C \cap N(m, \delta)$ there exists a connected subset P of $C \cap N(m, \epsilon/2)$ with p and m in P. Then by Theorem 4.4 there is a subcontinuum P' of $\bar{P}$ irreducible between p and m. Let q be another element of $C \cap N(m, \delta)$. By the same reasoning as above there is a continuum Q in $C \cap N(m, \epsilon/2)$, containing q and m, and there is an irreducible subcontinuum Q' of Q between q and m. Thus $P' \cup Q'$ is a subcontinuum of $C \cap N(m, \epsilon)$ containing p and q and so the arc $\overline{pq}$ (in the sense of Zoretti) is also contained in $N(m, \epsilon)$. $\square$

Note that unless the continuum C of the above proposition is irreducible we cannot guarantee that the arc between two of its points is unique, and as a result we can get no handle on the size of that arc. In fact, without the hypothesis of irreducibility, the proposition fails.

For example, consider the following set:



This is the topologist's sine wave augmented by constructing a ruler function on the y-axis from -1 to 1. The augmented set is connected im kleinen, but, for points m on the y-axis, there is no unique irreducible continuum connecting m with any other point p of the continuum. What is worse, there is always an irreducible subcontinuum of diameter 2 containing m and any point p of the continuum not on the y-axis.

As we have noted, Zoretti abandoned the definition of absolutely closed irreducible continuum in favor of the equivalent notion of a simple irreducible continuum. He was intent on proving a particular point -- that his "baby", the irreducible continuum, was the most natural mathematical counterpart to the intuitive notion of a line. The singlemindedness of his point of view apparently blinded him to value of his earlier definition.

The uncompromisingly harsh criticism he received from L.E.J. Brouwer for the mistakes in his 1909 paper seems to have discouraged Zoretti from publishing his research. In any case it is a fact that, after 1912, Zoretti published virtually no new journal articles, but rather spent his time in writing several texts -- not in topology, or point set theory, but in the more applied areas of mathematics. His interests seemed to focus on teaching.

It was during a set of lectures on analytic continuation given at the Collège de France in 1908-1909 that Zoretti first publicly introduced the concept of an irreducible continuum. At least one of the students attending those lectures was greatly impressed by their content. In the preface to his thesis [32], in 1911, Z. Janiszewski gave explicit credit to Zoretti's "beautiful lessons" for inspiring him to take up the torch and do his research on the theory of irreducible continua.

In this thesis, Janiszewski introduced the following new concepts. A continuum K is called a <u>continuum of condensation</u> of the continuum C if $\overline{C - K} = C$. This gives rise to two types of points in C. The points of <u>first type</u> are those which belong to no continuum of condensation of C. Those of <u>second type</u> are the points of C which belong to at least one continuum of condensation of C. The points of first type are further separated into two classes. Class I consists of those points p of first type for which some neighborhood of p contains no points of second type. Class II consists of all those points of first type which are not in class I.

The following example in E^2 illustrates these definitions (see Figure 24). Let C_n be a circle with center at the origin and radius $1/n$. Let S_n be the set (in polar coordinates) $\{(r,\theta):r = 1/(n+1) + (\pi/\theta)(1/(n(n+1))),\ \theta \geq \pi\}$. Let $C = (\bigcup_{n=2}^{\infty} C_n) \cup (\bigcup_{n=1}^{\infty} S_n) \cup \{(0,0)\}$. Let p be the point $(1,\pi)$. Then C is an

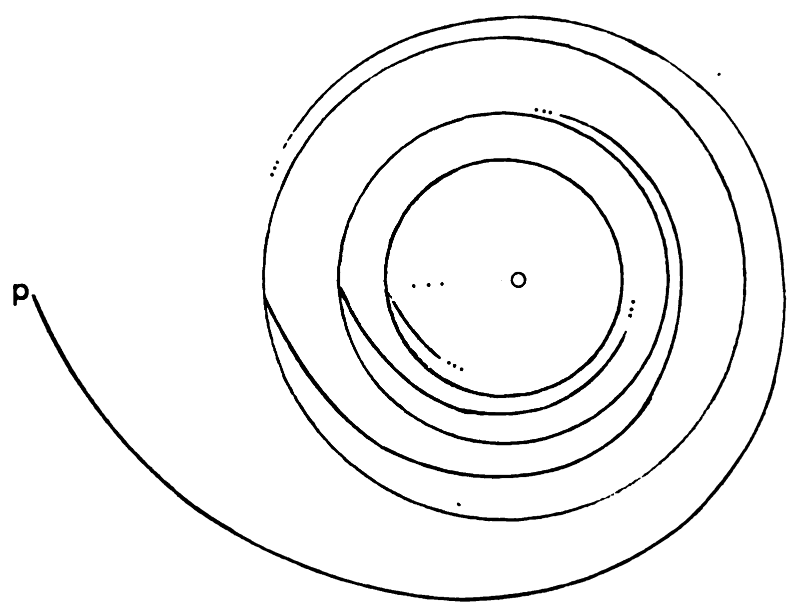


Figure 24

irreducible continuum between p and the origin. Further, point p is in class I and the origin is in class II. Each of the circles C_n, for $n \geq 2$, is a continuum of condensation of C. Thus, for example, the point $(\frac{1}{2},\pi/2)$ is of second type.

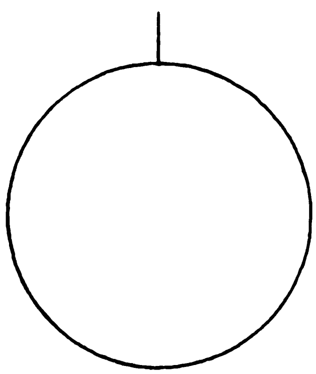
As Hahn later proved, the points of first type in a compact irreducible continuum are those, and only those, at which the continuum is connected im kleinen (see [25, p.219]). This is not true of an arbitrary continuum. In the unit square Q, every point is of second type, yet Q is a connected im kleinen continuum. A less trivial example is the union M of the following subsets of the plane:

$$M_x = \{(x,0): 0 \leq x \leq 1\},$$
$$M_0 = \{(0,y): 0 < y \leq 1\},$$
$$M_n = \{(1/n,y): 0 < y \leq 1\},\ n = 1,2,3,\cdots.$$

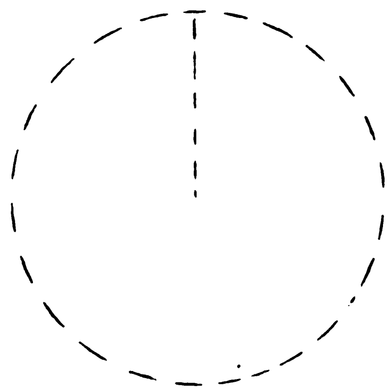
The point $(0,0)$ is a point of second type, since it belongs to the continuum of condensation $\bar{M}_0$ of M, but M is connected im kleinen at $(0,0)$.

In the same year, but shortly after Janiszewski finished his thesis, a little-known Italian mathematician named Pia Nalli published a paper [61] in which he defined a property equivalent to connectedness im kleinen in a different, though somewhat more general, setting than that of Janiszewski's work. Nalli was concerned primarily with the question: "What plane point sets are simple closed curves?" This led him to an examination of the boundaries of plane domains.

It should be noted that to Nalli a domain is a bounded, closed, connected point set with nonempty interior and such that each of its points is a limit point of interior points of the set. This definition excludes such sets as the following:



or



where in the second set the dashed line is the boundary of the set, not contained in the set.

Nalli's property, which he called condition (c) (presumably for "continuous curve" since there is no mention of any conditions (a), (b), (d), (e), etc.), is as follows. Let Δ be a closed, bounded point set in E^2 such that each of its points is a limit point of interior points of Δ. (Note that Δ need not be connected.) Let F be the frontier of Δ. F is said to satisfy <u>condition (c)</u> if, given any point a of F and any neighborhood S of this point, one can find a neighborhood S' of a, contained in S, such that for any point b of $F \cap S'$ and any $\epsilon > 0$, it is possible to construct an ϵ-chain in $F \cap S$ between a and b.

Although Nalli was concerned only with the boundary F of Δ, condition (c) automatically holds in the interior of Δ. Thus we can consider (c) to be a condition on the entire set Δ.

Unbeknown to Nalli, condition (c) is precisely what he needed to characterize continuous curves among compact, connected subsets of the plane. We prove this in the following proposition.

<u>Proposition 4.5</u>. Let F be a compact subset of the plane. F satisfies condition (c) iff F is connected im kleinen.

<u>Proof</u>: ($\Rightarrow$). Let a be an element of F and let N be a ball-neighborhood with center at a. Let O be a smaller ball-neighborhood centered at a. Let O' be the corresponding neighborhood of a given by condition (c). Let b be an element of $O' \cap F$. For every positive integer n, select a $1/2^n$-chain, $\{c_{n,i}\}_{i=1}^{k_n}$, in $O \cap F$ connecting a and b. Let S be the set of accumulation points of $M = \bigcup_{n=1}^{\infty} \left(\bigcup_{i=1}^{k_n} \{c_{n,i}\} \right)$.

Claim: S is a closed, connected subset of $\bar{O} \cap F$ containing a and b. That S is closed and contains a and b is clear. Suppose S is not connected. Then there are two nonempty closed subsets, E_1 and E_2, of $\bar{O} \cap F$, such that $E_1 \cap E_2 = \emptyset$ and $E_1 \cup E_2 = S$. Since E^2 is normal there are open subsets, O_1 and O_2, of E^2 such that $\bar{O}_1 \cap \bar{O}_2 = \emptyset$ and $E_1 \subset O_1$ and $E_2 \subset O_2$ (see [19, p.145]). Then $F - (O_1 \cup O_2)$ can contain at most a finite number of elements of M -- otherwise it would contain an accumulation point of M. Let δ_0 be the minimum of the distances of these points from $\bar{O}_1$ and $\bar{O}_2$, let $\delta' = d(\bar{O}_1, \bar{O}_2)$, and let $\delta = \min(\delta'/2, \delta_0/2)$. There exists an integer m such that $1/2^n < \delta$ for all $n \geq m$. Therefore, if a is in E_1, all but a finite number of elements of M must be contained in O_1. But this implies that E_2 is the empty set, since a finite set has no accumulation points, and we arrive at a contradiction. Hence S is connected. Consequently F is connected im kleinen at a.

($\Leftarrow$). Let a be an element of F, N any neighborhood of a and O a ball-neighborhood of a contained in N. Let O' be the neighborhood of a corresponding to O in the definition of connectedness im kleinen. Let b be an element of O' and let S be a connected subset of O containing a and b. Then by Theorem 4.1, for every positive number ϵ, there is a ϵ-chain in S between a and b. Since a is arbitrary, F satisfies condition (c).

Though Nalli defined condition (c) for E^2 only, the essential properties used in the proof of Proposition 4.5 hold in Euclidean n-space for any n. Therefore, we have generalized condition (c) to n-dimensions as follows: Let F be a bounded perfect point set in E^n. We say that F satisfies condition (c) if, for any point a in F and any neighborhood S of a, one can find a neighborhood S' of a, with $S' \subset S$, such that for any given point b of $F \cap S'$, and any $\epsilon > 0$, it is possible to construct an ϵ-chain in $F \cap S$ between a and b. The proof that this is equivalent to connectedness im kleinen is identical with that for $n = 2$.

The work of Nalli just discussed closely resembles some results of Arnaud Denjoy in a paper presented before the Paris Academy of Sciences in August, 1911 [16]. The question Denjoy was addressing was this: "What are the special properties the plane owes to its Analysis situs -- those which in particular, are not possessed by the torus or the Moebius strip?"

He claimed these properties can be taken to be orientability and what he called "biconnectivity." As a matter of historical interest the definition is included here. According to Denjoy, a closed subset M of Euclidean n-space is biconnected if it is connected and if, in addition, given any four points, $A, B, C,$ and D, of the set and any four continua in M, $(AB), (BC), (CD),$ and (DA), containing the indicated pairs of points, it is possible, for each fixed positive number ε, to choose a finite number of points of M, $M_{i,p}$ $(i = 0,1,2,\ldots,n;\ p = 0,1,2,\ldots,m)$, such that

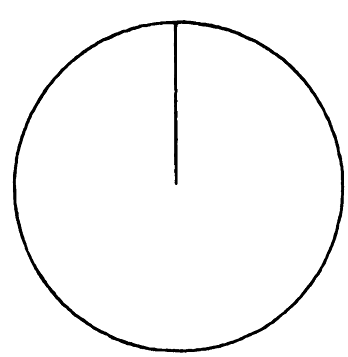
i. $M_{o,o} = A$, $M_{o,n} = B$, $M_{m,n} = C$, $M_{m,o} = D$; $M_{o,p} \in (AB)$, $M_{i,n} \in (BC)$, $M_{m,p} \in (CD)$, $M_{i,o} \in (DA)$; and

ii. $d(M_{i,p}, M_{i,p+1})$ and $d(M_{i,p}, M_{i+1,p})$ are less than ε for all i,p.

For open subsets of E^n is is further required that for each point $M_{i,p}$ there be a ball entirely contained in M of radius less than ε with center at $M_{i,p}$ and containing the points $M_{i,p+1}$ and $M_{i+1,p}$.

As can be seen, this seems to be a particularly cumbersome form of simple connectivity (in two-manifolds, at least). Perhaps it is just as well that it has not survived the test of time and indeed has relinquished its very name to another concept (see [46, p.135]).

Of greater interest to the present work are the following definitions given by Denjoy. An open subset M of E^n is said to be uniconnected if, for every pair of points, A and B, in M and for every given positive number ε, there exists a finite chain of points M_i $(i = 0,1,2,\ldots,n)$, such that $M_0 = A$, $M_n = B$, and for all i there is a ball $B_i \subset M$ with radius $< \varepsilon$ and center M_i and containing M_{i+1}. M is said to be uniformly uniconnected at a point P of its boundary if it is uniconnected and for every $\varepsilon > 0$, there exist $\delta_1, \delta_2 > 0$, such that whenever A and B are within δ_1 of P, there is a δ_2-chain in M between A and B with $\operatorname{diam}(\bigcup_{i=0}^{n} B_i) < \varepsilon$, where the B_i's are as given in the definition of uniconnectedness. The figure below is an example of a region which is uniconnected but not uniformly uniconnected:



Note that uniconnectedness is equivalent to path connectedness. Thus an alternative definition of uniform uniconnectedness is the following: An open subset M

of E^n is uniformly uniconnected at a point P of its boundary if, for every $\epsilon > 0$, there is a $\delta > 0$, such that, whenever A and B are in $N(P, \delta) \cap M$, there is a path from A to B contained in $N(P, \epsilon) \cap M$.

Using these definitions Denjoy established the following result.

Theorem 4.6. A bounded, uniformly uniconnected and simply connected planar region has as its boundary a simple closed curve.

Proof: See [16, p.426].

This theorem was rediscovered in 1918 by R.L. Moore [55], who also proved the necessity of the conditions. More will be said about this in Chapter VII.

L.E.J. Brouwer ([8], also in 1911) took a similar approach to that of Denjoy when he gave a definition of "unshieldedness" (Unbewalltheit). Brouwer was studying Jordan manifolds -- manifolds in E^n which are homeomorphic to the standard (n-1)-sphere S^{n-1}. Let M be a closed point set in E^n, and let G be one of the complementary regions determined by M. Then M is said to be unshielded from G if, for any accessible point q of the boundary of G and any two sequences $\{q_i\}_{i=1}^{\infty}$ and $\{q_i'\}_{i=1}^{\infty}$ contained in the boundary of G, all the points of which are accessible from G, and each of which converges to q, the path distance in G between q_i and q_i' tends to zero as i tends to ∞. Consider the modified Warsaw circle T of Figure 25. In this figure, the point $q = (0,-1)$ and the sequences $\{q_i = (\frac{2}{4i-1},-1)\}_{i=1}^{\infty}$ and $\{q_i' = (0,-1-1/n)\}_{i=1}^{\infty}$ illustrate that T is not unshielded from the region G (its exterior). However, T is unshielded from its interior region I. T is, of course, not a Jordan manifold.

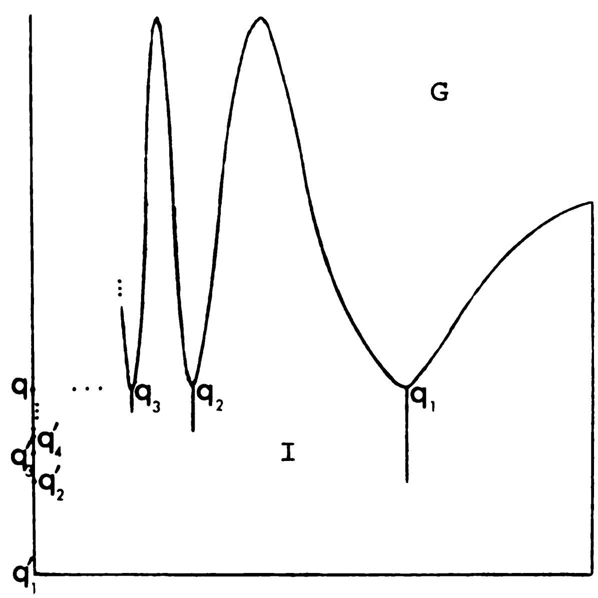


Figure 25

Brouwer was able to prove the following strong result.

Theorem 4.7. A Jordan manifold is unshielded from both its interior and its exterior regions.

Proof: See [8, p.322].

In the form in which it has been defined above, unshieldedness is not equivalent to uniform uniconnectedness, as the example of page 92 illustrates. However if we modify the definition by allowing the sequences $\{q_i\}$ and $\{q_i'\}$ to belong possibly to G, we can show that this modified version of unshieldedness is equivalent to uniform uniconnectedness.

Our proof requires the following lemmas.

Lemma 4.8. If $\{q_i\}_{i=1}^{\infty}$ $(\subset E^n)$ converges to q_o, then for every sequence of positive numbers $\{\delta_i\}_{i=1}^{\infty}$ converging monotonically to zero, there exists a subsequence $\{q_{n_i}\}_{i=1}^{\infty}$ of $\{q_i\}$ such that for all i, $d(q_{n_{i+1}}, q_o) < d(q_{n_i}, q_o) < \delta_i$ and $d(q_{n_i}, q_{n_j}) < \delta_i$ for all $j \geq i$.

Proof: For all $i = 1, 2, \ldots$ let $B_i = N(q_o, \delta_i/2)$. Let q_{n_1} be the first element of $\{q_i\}$ contained in B_1. Suppose $q_{n_1}, q_{n_2}, \ldots, q_{n_k}$ have already been chosen. Let $q_{n_{k+1}}$ be the first element of $\{q_i\}$ contained in B_{k+1} and such that $n_{k+1} > n_k$. The subsequence $\{q_{n_i}\}$ thus constructed is the desired subsequence. □

Lemma 4.9. Let q_o be a point of the boundary of a region G (in E^n). If G is uniformly uniconnected at q_o, then q_o is accessible from G.

Proof: Let $\{\epsilon_i\}_{i=1}^{\infty}$ be a sequence of positive numbers which converges monotonically to zero. Let δ_i correspond to ϵ_i in the definition of uniform uniconnectedness. Then there is a subsequence $\{\delta_i'\}_{i=1}^{\infty}$ of $\{\delta_i\}$ such that $\{\delta_i'\}$ converges monotonically to zero. Let $\{q_j\}$ be a sequence of points of G converging to q_o and let $\{q_{n_i}\}$ be a subsequence of $\{q_j\}$ which corresponds to $\{\delta_i'\}$ as in Lemma 4.8. Now construct a δ_i'-chain from q_{n_i} to $q_{n_{i+1}}$ and contained in $N(q_o, \epsilon_i)$. Since according to the definition of uniconnectedness there exists such a chain which is completely contained in a connected open subset of G, there is a path ℓ_i in G from q_{n_i} to $q_{n_{i+1}}$ contained in $N(q_o, \epsilon_i)$. Thus $\bigcup_{i=1}^{\infty} \ell_i$ is a path in G from q_{n_1} to q_o, and so q_o is accessible from G. □

Proposition 4.10. If p is a point of the boundary γ of a region G in E^n, then γ is unshielded from G at p (in the modified sense) iff G is uniformly uniconnected at p.

<u>Proof</u>: ($\Rightarrow$). Suppose that γ is unshielded from G at p, and suppose also that G is not uniformly uniconnected at p. Let $\delta_i = 1/i$ for $i = 1,2,\cdots$. Then there exists $\varepsilon > 0$ such that, for each δ_i, there is a pair of distinct points, q_i and q_i', in $N(p,\delta_i)$, which cannot be joined in G by a path lying entirely in $N(p,\varepsilon)$. That is, the path distance in G from q_i to q_i' is $\geq \varepsilon$. But the existence of the sequences $\{q_i\}_{i=1}^{\infty}$ and $\{q_i'\}_{i=1}^{\infty}$ contradicts the supposition that γ is unshielded from G at p.

($\Leftarrow$). Suppose that G is uniformly uniconnected at p. By Lemma 4.8, p is accessible from G. Let $\{q_i\}$ and $\{q_i'\}$ be two sequences of points in $\bar{G}$ each of which converges to p.

Let $\varepsilon > 0$ be given, and let δ be the positive number which corresponds to ε in the definition of uniform uniconnectedness. Then there exist two positive integers, N and N', such that for all $n \geq N$ and $n' \geq N'$, q_n and $q_{n'}$ are both contained in $N(p,\delta)$. Let $N_o = \max(N,N')$, and let n be an integer greater than N_o.

<u>Case 1</u>: q_n and $q_{n'}$ are contained in G.

Since both q_n and $q_{n'}$ are in $N(p, \delta)$, there is a path in G connecting q_n and $q_{n'}$ which is contained in $N(p, \epsilon)$. Thus the path distance from q_n to $q_{n'}$ relative to G is less than ϵ.

Case 2: One of q_n or $q_{n'}$ is contained in Υ (say q_n), while the other (say $q_{n'}$) is contained in G.

Let $\rho = d(q_n, p)$, and let $r = \delta - \rho$. Let a be a point in $N(q_n, r)$ such that there is a path ℓ_1 from a to q_n contained in $N(q_n, r)$. a is in $N(p, \delta)$ as is $q_{n'}$, so there is a path ℓ_2 from a to $q_{n'}$ contained in $N(p, \epsilon)$. Thus $\ell_1 \cup \ell_2$ is a path in $N(p, \epsilon)$ from q_n to $q_{n'}$.

Case 3: Both q_n and $q_{n'}$ are elements of Υ.

By an argument similar to that of case 2 we can construct a path from q_n to $q_{n'}$ contained in $N(p, \epsilon)$. Since ϵ is arbitrary, the path distance from q_n to $q_{n'}$ tends to zero as n tends to ∞, and, thus, Υ is unshielded from G at p. $\square$

It is interesting that in every neighborhood of a point of uniform uniconnectedness there may exist points at which the set is not uniformly uniconnected. In Figure 25, the sequence $\{q_i\}$ converging to q is an example of this phenomenon.

Despite this example one cannot help but notice the similarity between uniform uniconnectedness (unshieldedness) and uniform connectedness im kleinen. With the following proposition we have made the relationship of these two concepts precise.

Proposition 4.11. A bounded region G is uniformly connected im kleinen iff it is uniformly uniconnected at each point of its boundary γ.

Proof: ($\Rightarrow$). Suppose G is uniformly connected im kleinen, and let p be a point of γ. Let $\varepsilon > 0$ be given and let $\delta > 0$ be the number corresponding to ε in the definition of uniform connectedness im kleinen. Let q and q' be points in $G \cap N(p, \delta/2)$. Then $d(q,q') < \delta$. Therefore, there exists a path ℓ from q to q' with $\operatorname{diam}(\ell) < \varepsilon$. Thus, G is uniformly uniconnected at p.

($\Leftarrow$). Suppose G is uniformly uniconnected at each point of γ. Let $\varepsilon > 0$ be given, and, for each point $p \in \gamma$, let δ_p ($< \varepsilon$) correspond to ε in the definition of uniform uniconnectedness. Since γ is compact, $N(p, \delta_p/4)$ can be reduced to a finite subcover $N(p_i, \delta_{p_i}/4)$ of γ. Let $\delta' = d(\gamma, \mathcal{C}(\bigcup_{i=1}^{k} N(p_i, \delta_{p_i}/4)))$ and let $\delta = \min\{\delta'/4, \delta_{p_1}/4, \ldots, \delta_{p_k}/4\}$. (Note that $\delta < \varepsilon/2$.)

Let q and q' be two points of G with $d(q,q') < \delta$. It will suffice to show that there exists a path ℓ in G from q to q' with $\operatorname{diam}(\ell) < \varepsilon$.

<u>Case 1</u>: $d(q,\gamma) < \delta$ or $d(q',\gamma) < \delta$. Then q and q' are both in $N(p_j, \delta_{p_j}/2)$ for some j. So, by the uniform uniconnectedness at p_j, there is a path ℓ_j from q to q' with $\operatorname{diam}(\ell_j) < \varepsilon$.

<u>Case 2</u>: $d(q,\gamma) > \delta$ and $d(q',\gamma) > \delta$. Then q and q' are in $N(q,\delta) \subset G$ and hence the straight line segment $\overline{qq'} \subset N(q,\delta)$ and $\operatorname{diam}(\overline{qq'}) < \varepsilon$. $\square$

As Theorems 4.6 and 4.7 indicate, Denjoy and Brouwer were significantly closer to the eventual point set characterization of "curve" than were their contemporaries. But, even so, they were somewhat limited in their perspective. The reason Denjoy discarded this line of research might have been summarized in a statement he made in 1955:

> "I have always found repugnant the purely formal generalization, an easy and sterile exercise if it does not lead to the solution of some problem already existing and different from that which gave birth to the first concept." [18, p.vii]

As we shall see (Chapter VI) the approach which Nalli, Denjoy and Brouwer took -- that is, the study of boundaries of regions -- was a very natural one from the point of view of complex analysis.

CHAPTER V

THE HAHN-MAZURKIEWICZ THEOREM

Stefan Mazurkiewicz' mathematical career began about three years before he obtained his Ph.D. In 1910 he published a note in Compte Rendu, Paris, under the name Étienne Mazurkiewicz, giving a topological proof of Theorem 4.4 (see [48, p.296]), which had earlier been proved by Janiszewski using a method involving transfinite numbers. This first paper signals the early contact that Mazurkiewicz had with Zoretti's work on irreducible continua.

Mazurkiewicz then went on to obtain his doctorate in 1913 under Waclaw Sierpiński at the University of Lwow, writing his thesis on characterizations of space-filling curves. Also in 1913, he published two articles ([49] and [50], originally in Polish, but later translated into French) characterizing Jordan curves, or what later came to be known as Peano spaces. These articles seemed to attract no attention until Mazurkiewicz published an enlarged paper, Sur les lignes de Jordan [51], in the first volume of Fundamenta Mathematicae (1920) -- of which he was one of the co-founders. This article was an expansion of the two shorter papers into a more comprehensive study.

The original version of Mazurkiewicz' characterization of Peano spaces was given for n-dimensional Euclidean space. However, in [51] he observed that the greatest part of the reasoning can be applied, with little or no alteration of the proofs, to "suitable" abstract spaces. For a space to be suitable he required that it be metric. This is no real restriction, since, as it turns out, every Peano space is metrizable (see, for example, [26, p.201]).

Let us examine Mazurkiewicz' characterization of Peano spaces in closer detail. We give here the 1913 version of Mazurkiewicz' theorem:

<u>Theorem 5.1</u>. A continuum $\Gamma \subset E^n$ is a Jordan curve iff

(i) Γ is bounded;

(ii) for every pair of points, A and B of Γ, there exists a simple arc (AB) (in the usual sense) contained in Γ joining the two points; and

(iii) the diameter of (AB) tends toward zero as the point B tends toward A.

(Condition (iii) is later replaced by the equivalent condition (iii'): "the diameter of (AB) tends uniformly toward zero with the distance between the points A and B.")

Proof: See [49, p.305].

We note that conditions (ii) and (iii) are merely a generalization to arbitrary continua of the concept of an absolutely closed irreducible continuum, introduced by Zoretti in 1909 (see page 78). It is a testimony to the creative genius of Mazurkiewicz that he was able to recognize the value of this concept. Zoretti himself seems to have abandoned it in favor of the idea of a simple irreducible continuum -- a concept which, although equivalent, led him into a completely different vein of thought.

In the second of his 1913 papers, Mazurkiewicz gives the following refinement of Theorem 5.1:

Theorem 5.2. A bounded continuum $\Gamma \subset E^n$ is a Jordan curve if to every pair of points $a,b \in \Gamma$ there corresponds a continuum $C(a,b) \subset \Gamma$ containing a and b, such that the diameter of $C(a,b)$ tends toward zero with the distance between the points a and b.

Proof: See [50, p.941].

The refinement here is only slight, amounting to a replacement of the requirement of a simple arc between a and b by that of the existence of a subcontinuum containing a and b, but it is one step toward greater generality. Note also that Theorem 5.2 proves only the sufficiency of the condition for Γ to be a Jordan curve. The necessity of this condition is a trivial consequence of Theorem 5.1. That is, the simple arc (ab) of Theorem 5.1 between a and b is just such a subcontinuum $C(a,b)$.

In his presentation in Sur les lignes de Jordan, Mazurkiewicz introduced the concept of "genre" of a point. A point of a continuum C is said to be of first genre if the "oscillation" of the continuum at that point is zero. The oscillation $\sigma(p)$ of a continuum C at a point p is defined by

$$\sigma(p) = \lim_{x,y \to p} \left(\inf_{C(x,y)} \operatorname{diam}(C(x,y)) \right)$$

where $C(x,y)$ denotes any subcontinuum of C containing x and y. A point p with $\sigma(p) > 0$ is said to be of second genre.

The idea of genre of a point gives greater mathematical precision to the condition expressed in Theorem 5.2. Thus Mazurkiewicz was able to further simplify the statement of his theorem.

<u>Theorem 5.3</u>. A necessary and sufficient condition for a bounded continuum A to be a Jordan curve is that all of the points of A be of first genre.

<u>Proof</u>: See [51, p.191].

He then went on to define a generalized continuous curve (one where the possibilities for its domain include not only closed intervals, but also half lines), and finally a generalized Jordan curve -- "any continuum, bounded or not, formed solely of points of first genre." Hence came Mazurkiewicz' final modification of the theorem:

<u>Theorem 5.4</u>. Every generalized Jordan curve A is the image of a generalized continuous curve.

<u>Proof</u>: See [51, p.193].

Note that the converse of Theorem 5.4 is false as the following example (due to Mazurkiewicz) clearly shows. Let the curve $A \subset E^2$ be defined as the union of the sets

$$\{(0,1-t):0 \leq t \leq 1\},$$
$$\{(t-1,0):1 < t \leq 2\},$$
$$\{(2/t,\sin^2 t):2 < t\}.$$

(See Figure 26.)

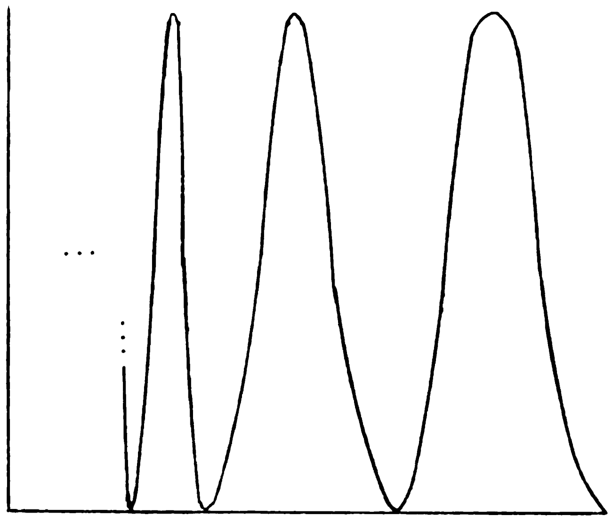


Figure 26

We have observed that conditions (ii) and (iii) of Theorem 5.1 are almost identical with the definition of absolutely closed irreducible continuum given by Zoretti four years earlier. Why then does the credit for the concept of connectedness im kleinen go to Mazurkiewicz while Zoretti remains an almost forgotten figure? Clearly Mazurkiewicz made the notion precise and put it in a more general setting, but more to the point, he used it. As is true with so much of mathematics, Mazurkiewicz built on the work of his predecessors, Zoretti and Janiszewski, and perhaps because of his particular training and the unique insight he was able to bring to bear on the problems involved in characterizing Jordan curves, he brought the work of his predecessors into clearer focus. His was certainly the first characterization of Jordan curves to be published.

Theorems 5.1, 5.2, and 5.3 provide us with a rough outline of the refinement of the concept of connectedness im kleinen as Mazurkiewicz developed it over a period of seven years. The actual change in the concept during that period seems almost insignificant. The process appears rather to be one of focusing in on the essence of the underlying idea. The final product was the definition of a point of first genre, or point of connectedness im kleinen.

The first papers of Hans Hahn that dealt with connectedness im kleinen, appeared in print in 1914. In 1913 he had read a paper to the Versammlung deutscher Naturforscher und Ärzte at Vienna, Austria, entitled, Über die allgemeinste ebene Punktmenge, die stetiges Bild einer Strecke ist [21]. In this paper he presented four necessary and sufficient conditions for a plane point set M to be the continuous image of a line segment. The first three conditions were already known to be necessary but not sufficient. They were

1. The set M is bounded;
2. The set M is closed;
3. The set M is connected.

Hahn then added a fourth condition so that the four conditions together were necessary and sufficient. It is here we find the first definition of connected im kleinen as such. Let p be a point of the set M. M is said to be connected im kleinen at p if "to each positive number ϵ there corresponds a positive number δ such that for every point p' of M lying in the δ-neighborhood of p there is a closed and connected subset of M containing both p and p' and which lies entirely in the ϵ-neighborhood of p." (Recall that a set is connected im kleinen if it is connected im kleinen at each of its points.) Thus the fourth condition is

4. The set M is connected im kleinen.

In the process of proving the sufficiency of these four conditions, Hahn introduced an interesting concept that may be called a local component of a set. Given a subset M of a metric space, a point p of M and a positive real number r, the local component of M at p with respect to r is the set of all points x of M for which there exists a subcontinuum of M containing both x and p and contained in $\overline{N(p,r)}$, together with all the limit points in M of such points. We denote this set by $M^*(p,r)$.

It is interesting to note that, in constructing a mapping from the closed unit interval onto a given bounded, connected im kleinen continuum M in E^2, Hahn followed the pattern of Hilbert's construction of a space-filling curve. He first selected an arbitrary sequence of positive real numbers r_i having the property that $\lim_{i\to\infty} r_i = 0$. Then he covered M with the sets $M^*(p,r_1)$, for all points p in M.

In the process of the construction, Hahn proved the following intermediate results.

<u>Lemma 5.5</u>. Each $M^*(p,r)$ is connected.

<u>Proof</u>: See [22, p.2442].

<u>Lemma 5.6</u>. Each $M^*(p,r)$ is connected im kleinen.

<u>Proof</u>: See [22, p.2443].

<u>Lemma 5.7</u>. For each point p of M and each positive number r, there is an open neighborhood U of p with U contained in $M^*(p,r)$.

<u>Proof</u>: See [22, p.2441].

<u>Lemma 5.8</u>. If the bounded continuum M $(\subset E^n)$ is connected im kleinen, then every pair of points in M can be joined by a path contained in M.

Proof: See [22, p.2436].

By Lemma 5.7, such a covering by sets $M^*(p,r)$ induces a covering by open sets and hence the Heine-Borel Theorem can be invoked to produce a finite subcovering $M_1^{(1)} = M^*(p_{1,1},r_1)$, $M_2^{(1)} = M^*(p_{1,2},r_1),\ldots,$ $M_{n_1}^{(1)} = M^*(p_{1,n},r_1)$. The next step in the construction is to establish the fact that, given any two points p and q of M, the collection $\{M_i^{(1)}\}_{i=1}^{n_1}$ can be ordered in such a way that the following conditions are satisfied:

(i) p is contained in $M_1^{(1)}$;

(ii) q is contained in $M_{n_1}^{(1)}$; and

(iii) $M_i^{(1)} \cap M_{i+1}^{(1)} \neq \emptyset$, for $i = 1,2,\ldots,n_1-1$.

Using this ordering, Lemmas 5.5, 5.6, and 5.7 can be applied to give a first approximation to the desired "space-filling" curve. For the next stage of construction, M is replaced in turn by each of the $M_i^{(1)}$'s, and the above process is repeated, using r_2 in place of r_1. The induction step is now obvious.

Hahn proved that this process does indeed give a continuous mapping from the closed unit interval to the set M [22, p.2439]. Thus the sufficiency of conditions 1-4 is proved for the plane.

Surprising as it may seem, this proof does not carry over directly to higher dimensions, since, as is shown by the following example given by Hahn, $M^*(p,r)$ may fail to be connected im kleinen (see Figure 27). Let α be a plane in E^3. Let O be a point of α, and let g be a straight line in α passing through O. Now using O as a center, construct a circle K_o of radius 1, and circles K_n and K_n' of radii $\frac{n}{n+1}$ and $\frac{n+1}{n}$ respectively, for $n = 1,2,\cdots$. Now, for $n \geq 1$, rotate the circles K_n and K_n', in the same direction about the line g, through angle $\theta_n = \pi/2n$. The set of points through which the circumferences of the circles K_n and K_n' pass will be denoted by M_n. Let N_n denote the annulus between the images of K_n and K_n'. Let G be the subset of g which forms a diameter of the circle K_o. Now let M be the union of all of the above described sets, K_i, K_i', M_i, N_i, and G.

In this example the set $M^*(O,1)$ is not connected im kleinen. (A cross-section of $M^*(O,1)$ is given in Figure 28.) That is, for all points of the circle K_o, except the two points where g intersects K_o, $M^*(O,1)$ is not connected im kleinen.

To overcome this difficulty, Hahn replaced the sets $M^*(p,r)$ used in the two-dimensional proof by the

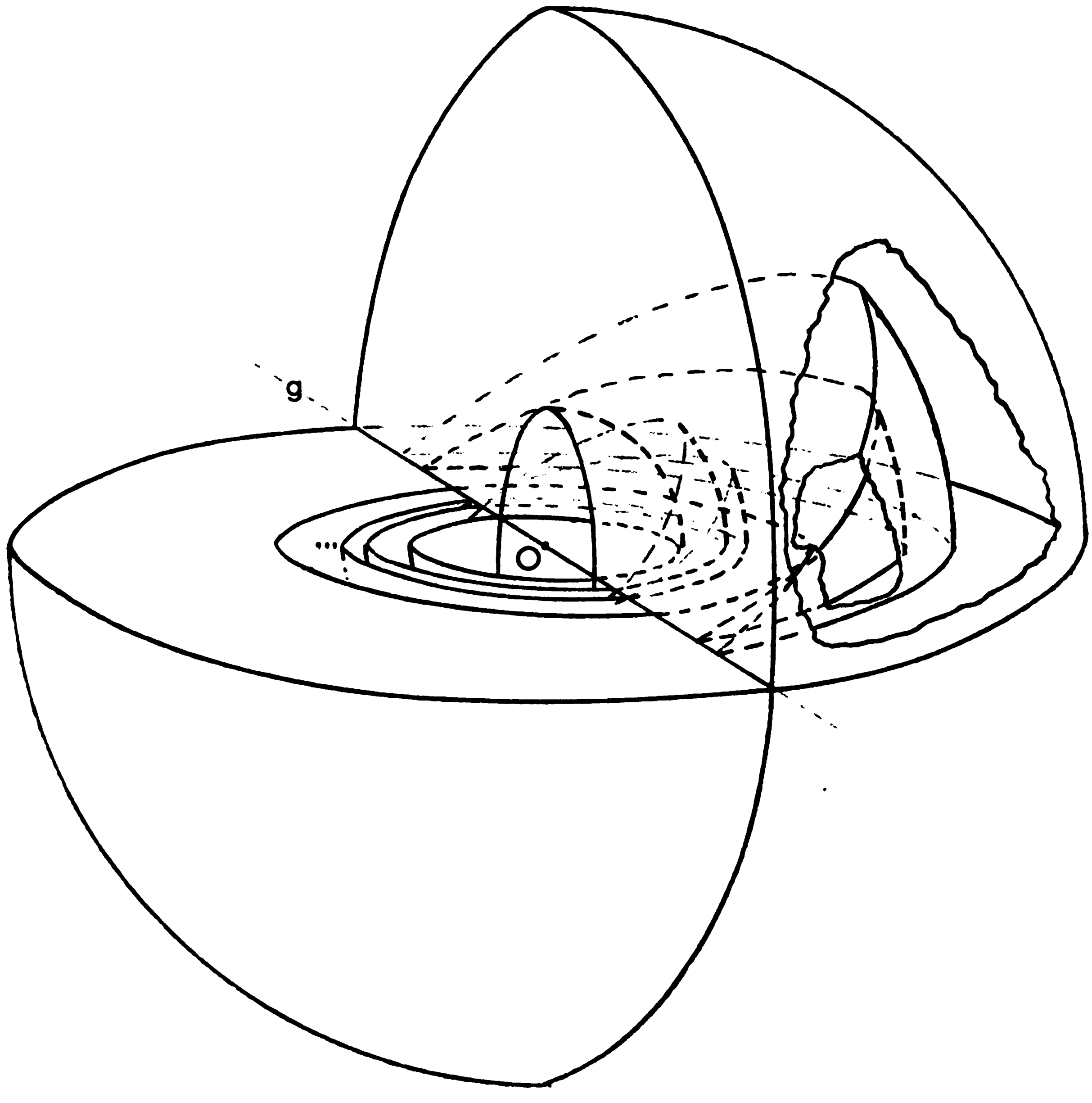


Figure 27

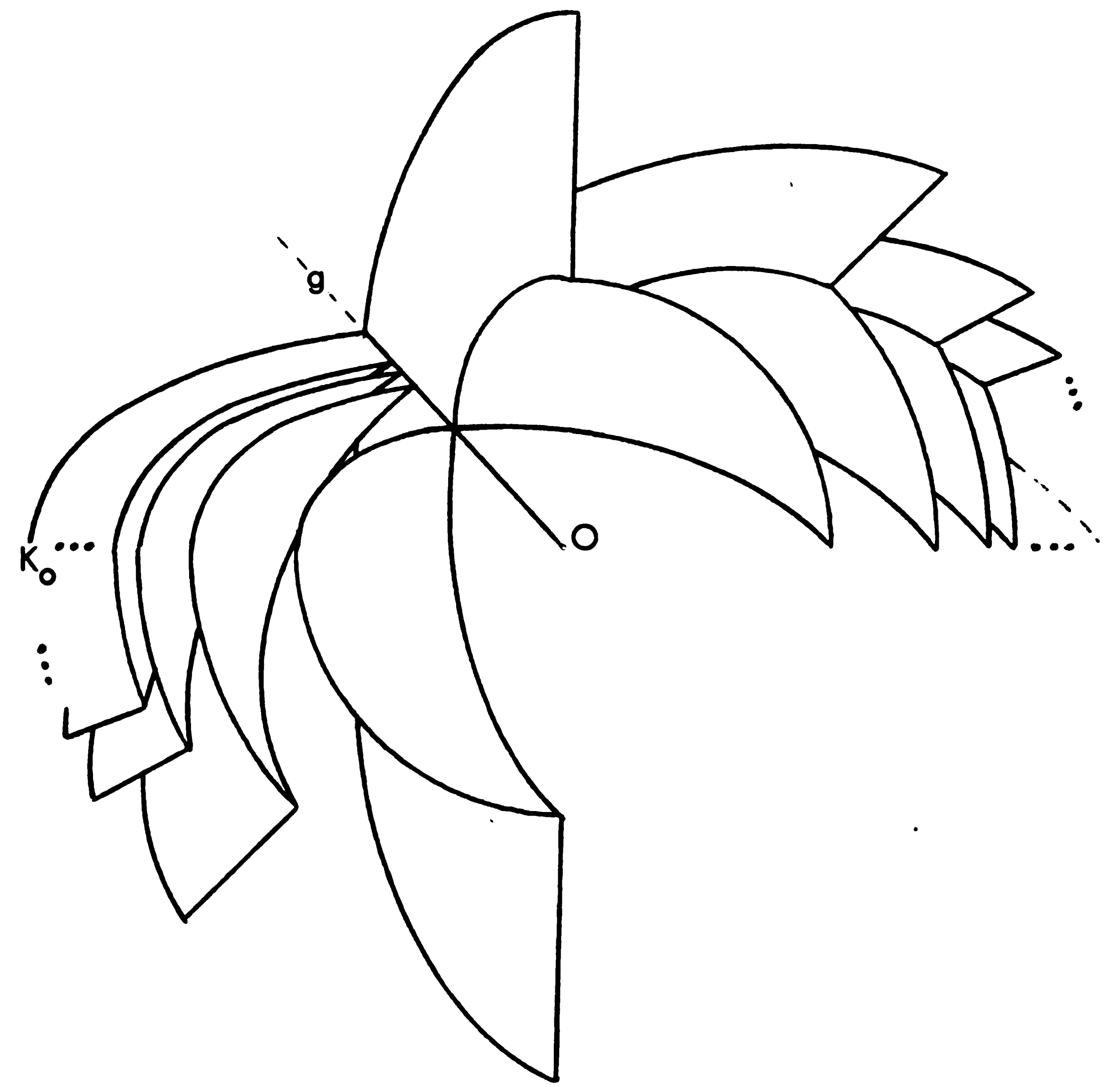


Figure 28

sets $M^{**}(p,r)$ described below. Let p_o be a fixed element of M and let r be a given positive number. $M^{**}(p_o,r)$ is obtained from $M^*(p_o,r)$ by first choosing any positive number $r_o' < r$, and considering $M^*(p_o,r_o')$. Since $r - r_o > 0$, a positive number r_1 can be selected with $r_o + r_1 < r$. Cover $M^*(p_o,r_o)$ with the collection $\{M^*(p,r_1)\}_{p \in M^*(p_o,r_o)}$, and, with the aid of Lemma 5.7 and the Heine-Borel Theorem, reduce this covering to the finite subcovering $\{M^*(p_1{}^{(1)},r_1),\ldots, M^*(p_{n_1}{}^{(1)},r_1)$. (Note that to invoke Lemma 5.7, Hahn had to use the fact that the set M itself is connected im kleinen.) Again Hahn used the notation $M^*(p_j{}^{(1)},r_1) = M_j{}^{(1)}$. Now $M^*(p_o,r_o)$ is contained in the union of the sets $M_j{}^{(1)}$, and since $r_o + r_1 < r$, this union is contained in the set $M^*(p_o,r)$. If $M^{(1)} = \bigcup_{j=1}^{n_1} M_j{}^{(1)}$, then every point of $M^{(1)}$ is at most a distance $r_o + r_1$ from p_o. Now choose a positive number r_2' such that $r_o + r_1 + r_2' < r$.

Since the sets $M_i{}^{(1)}$ are closed, if two of these sets, say $M_i{}^{(1)}$ and $M_j{}^{(1)}$, are disjoint, the distance between them is positive. Let σ_1 denote the smallest of these distances, and let r_2'' be a positive number $\sigma_1/4$. Let r_2 denote the smaller of the two numbers r_2' and r_2''.

As before $M^{(1)}$ is covered with the sets $\{M^*(p,r_2)\}_{p\in M^{(1)}}$. Because of the choice of r_2, for each p in $M^{(1)}$, $M^*(p,r_2) \subset M^*(p_o,r)$. Further, if p' and p'' belong to different sets $M_i^{(1)}$ and $M_j^{(1)}$, then $M^*(p',r_2)$ and $M^*(p'',r_2)$ can have a point in common only if $M_i^{(1)}$ and $M_j^{(1)}$ have a point in common.

The covering of $M^{(1)}$ is next reduced to a finite subcovering. For each of the sets $M_i^{(1)}$, that part of this subcovering which covers it is denoted by

$$M_{i,1}^{(2)},\ldots,M_{i,n_2}^{(2)},$$

where without loss of generality n_2 is assumed to be the same for all i.

Following this pattern the sequence of subsets $M_{i_1 i_2,\ldots,i_\nu}^{(\nu)}$ $(i_1 = 1,2,\ldots,n_1;\ i_2 = 1,2,\ldots,n_2;\ldots;$ $i = 1,2,\ldots,n;\ \nu = 1,2,\ldots)$ of M is constructed inductively. Every point of each of these subsets is at a distance from p_o at most $r_o + r_1 + r_2 +\ldots+ r_\nu < r$, and each of the sets satisfies the following properties:

(i) $M_{i_1,i_2,\ldots,i}^{(\nu)}$ lies entirely in a ball of radius r_ν, where $\lim\limits_{\nu\to\infty} r_\nu = 0$;

(ii) The two sets $M^{(\nu)}_{i'_1, i'_2, \ldots, i'}$ and $M^{(\nu)}_{i''_1, i''_2, \ldots, i''_{-1}, i''}$ have a point in common only if the sets $M^{(1)}_{i'_1}$ and $M^{(1)}_{i''_1}$, $M^{(2)}_{i'_1, i'_2}$ and $M^{(2)}_{i''_1, i''_2}, \ldots, M^{(\nu-1)}_{i'_1, i'_2, \ldots, i'_{\nu-1}}$ and $M^{(\nu-1)}_{i''_1, i''_2, \ldots, i''_{\nu-1}}$ also have a point in common.

Now let $M^{**}(p_o, r)$ be the closure of the union of all the sets $M^{(\nu)}_{i_1, i_2, \ldots, i_\nu}$. This union is clearly bounded (in the general setting, compact), and it can be easily shown that it is connected. Hahn also proved (via a very lengthy and complicated argument) that $M^{**}(p_o, r)$ is connected im kleinen.

Once these properties were established, the proof of the sufficiency of Hahn's four conditions carried over directly from the planar case to the general metric space with the substitution of $M^{**}(p,r)$ wherever $M^{*}(p,r)$ was used.

Hahn seems to have been unaware (at least through 1914) of the work of Zoretti, Janiszewski and Mazurkiewicz, on irreducible continua and points of connectedness im kleinen. In [22], which gives a "set theoretic characterization of continuous curves," Hahn does make one brief reference to an article by Zoretti appearing in the Encyclopédie des sciences mathématiques, volume II [93]. The theorem he cites is one giving conditions under which the limit set of a sequence of closed sets is connected. Even though Zoretti's article gives a brief description of irreducible continua and some of the results of Zoretti and Janiszewski, the direction of Hahn's research would indicate that he took little or no notice of it. This may not be surprising since the connection with connectedness im kleinen, which is quite apparent in Zoretti's earlier paper, was removed by repeated polishing.

CHAPTER VI

THE THEORY OF PRIME ENDS

The problem of mapping one simply connected region onto another was of great interest to the complex analysts of the late nineteenth and early twentieth centuries. A common approach to the study of such mappings was to first approximate the given mapping by a sequence of more elementary mappings which converge pointwise on the interior of the region to the given mapping. Unfortunately examples were known which have the property that they diverge at every point of the boundary (see, for example [12, p.144]). It was of particular interest then to determine what restrictions must be placed on the boundary of such a region to insure that a conformal mapping of its interior onto the interior of the circle $|z| = 1$ would induce a homeomorphism between the respective boundaries.

H.A. Schwartz showed [76] that this is always the case if the boundary of the region to be mapped onto the interior of the circle consists of a finite number of regular analytic arcs. Painlevé later showed [65] that it was sufficient to require that the boundary be piece-wise differentiable.

In 1900, W. Osgood [63, p.56] conjectured that if the boundary of a simply connected region were homeomorphic to a circle then any conformal mapping of the interior of the region onto the interior of the circle would induce a homeomorphism between the boundaries. This conjecture was finally proved by C. Carathéodory in 1912 [13] and was perhaps the motivation for the concepts which we discuss below.

At about the same time Hahn and Mazurkiewicz were developing connectedness im kleinen, Carathéodory published a paper [14] in which he defined the concept of a prime end of a bounded simply connected region. He first presented his work in September of 1911 at a meeting of natural scientists in Karlsruhe, Germany. At least part of the development parallels some work on conformal mappings of simply connected regions published by Eduard Study in 1912. Because a copy of Study's book has not been located, the present author can only guess at the precise nature of Study's work from the occasional references to it which Carathéodory makes.

Before defining a prime end, the following definitions are needed. By a <u>path</u> Carathéodory meant a finite, or countably infinite sequence of closed straight line segments $\ell_1, \ell_2, \ldots,$ or $\ell_0, \ell_{\pm 1}, \ell_{\pm 2}, \ldots,$ such that

(i) for every i, ℓ_i has exactly one point (an endpoint) in common with each of the segments ℓ_{i-1} and ℓ_{i+1}, and

(ii) for every pair of indices i, j with $i < j-1$, $\ell_i \cap \ell_j = \emptyset$.

The only exception to condition (ii) can occur in the finite case when the last segment ℓ_n may have a vertex in common with the first segment ℓ_1 to form a <u>closed path</u>. If the path consists of an infinite number of segments, it is further stipulated that

(iii) $\lim_{n \to \infty} (\mathrm{diam}(\ell_n)) = 0$, and

(iv) given any sequence of points, with only a finite number of them taken from any one segment of the path, at most one of the following occurs: (a) its limit points do not lie on the path; or (b) $i \geq 1$ for all i, and the unique limit point of the sequence is the vertex of ℓ_1 which is not contained in ℓ_2 -- giving rise to a closed path.

A limit point of a path not lying on the path will be called an endpoint of the path.

Let G be a bounded simply connected region in the plane. (Throughout this discussion, unless otherwise indicated, all regions and subregions will be assumed to be bounded, simply connected regions of the plane.) A path k, exactly one endpoint of which lies on the boundary γ of G, is called an end-cut of G. A path q, both of the endpoints of which lie on γ, will be called a cross-cut of G.

Let $q_1, q_2, \ldots,$ be a sequence of cross-cuts of G satisfying the two conditions:

(i) No two of the cross-cuts have any points in common. In particular, their endpoints are all distinct;

(ii) Each cross-cut q_n separates q_{n-1} from q_{n+1} in the sense that if P is a point of q_{n-1} and Q is a point of q_{n+1} and ℓ is any path in G connecting P and Q, then ℓ has at least one point in common with q_n.

Any sequence satisfying these two conditions will be called a chain of cross-cuts.

Corresponding to each chain of cross-cuts, one can define a chain of subregions $g_1, g_2, \ldots,$ such that q_i froms part of the boundary of g_i and $g_i \supset g_{i+1}$ for $i > 1$. Note that Carathéodory refers to such a sequence of regions as a chain only if it corresponds to a chain of cross-cuts.

The <u>end</u> $\mathcal{S}_g$ corresponding to the chain $g_1, g_2, \ldots,$ of subregions of G is defined by the following set of axioms.

<u>Axiom 6.1</u>: If H is an arbitrary region (not necessarily a subregion of G), then the end $\mathcal{S}_g$ is contained in H iff H contains g_n for some n.

<u>Axiom 6.2</u>: Given two chains of subregions of G, $g_1, g_2, \ldots,$ and $h_1, k_2, \ldots,$ with their respective ends, $\mathcal{S}_g$ and $\mathcal{S}_h$, $\mathcal{S}_h$ is contained in $\mathcal{S}_g$ iff $\mathcal{S}_h$ is contained in every g_n. In this case $\mathcal{S}_h$ is said to be a <u>subend</u> of $\mathcal{S}_g$.

<u>Axiom 6.3</u>: A sequence of points converges to an end $\mathcal{S}_g$ iff every region g_n contains all but a finite number of the points of the sequence.

<u>Axiom 6.4</u>: A point P is contained in the end $\mathcal{S}_g$ iff P is contained in the closure of each region g_n.

Axiom 6.5: Two ends $\mathcal{E}_g$ and $\mathcal{E}_h$, defined by the chains of subregions $g_1, g_2, \ldots,$ and $h_1, k_2, \ldots,$ respectively, are identical iff each is a subend of the other.

A prime end is an end which contains no proper subends. The prime end defined by a chain of subregions $g_1, g_2, \ldots,$ will be denoted by E_g.

The following are a few of the principal properties of prime ends which Carathéodory proved:

Theorem 6.6. Every infinite sequence of points $P_1, P_2, \ldots,$ of the region G contains a subsequence which converges to either an interior point of G or to a prime end of G. [14, p.341]

Theorem 6.7. Every prime end of a region can be defined by a chain of cross-cuts which lie on concentric circles and which converge to the common center of these circles. [14, p.343]

Theorem 6.8. A prime end of a region G contains no interior points of G. [14, p.344]

Theorem 6.9. Under any conformal mapping of the interior of a region G onto the interior of the circle $|z| = 1$, the prime ends of G correspond to the boundary points of the circle in a one-to-one fashion. Each sequence of points of G which converges to a prime end E_g corresponds to a sequence of points of $|z| < 1$ which converges to the image of E_g, and vice versa. [14, p.350]

A prime end is said to be of <u>first type</u> if it contains only a single point. For example, a disk has only prime ends of first type. Carathéodory gave the following characterization of such prime ends.

<u>Theorem 6.10</u>. A prime end E_g contains a single point ε iff the regions $g_1, g_2, \ldots,$ of any chain which defines E_g, converge to ε. The point ε is always an accessible point of E_g, that is, in the terminology of Carathéodory, ε is the endpoint of an end-cut of G which converges to E_g. [14, p.352]

The next result, though easily proved, is important in relating the theory of prime ends to connectedness im kleinen.

<u>Theorem 6.11</u>. A prime end E_g contains at most one accessible point. [14, p.353]

Theorem 6.11 implies that any prime end of G containing more than one point must contain non-accessible points. That being the case, the boundary of G is not connected im kleinen.

A point ϵ of a prime end E_g for which there exists a chain of cross-cuts $q_1, q_2, \ldots,$ defining E_g and such that the limit set of the sequence $\{q_i\}$ is $\{\epsilon\}$ is called a <u>major point</u> of E_g. A point of E_g which is not a major point is called a <u>minor point</u> of E_g. It follows immediately from Theorem 6.7 that every prime end contains at least one major point, but, as the example of the disk shows, a prime end need not contain any minor points. For an example of a prime end which does contain minor points, let G be the interior of the Warsaw circle T. Consider a sequence of concentric circular arcs in $G \cup T$ with center at the point $(0,-1)$. The cross-cuts determined by these circular arcs define the prime end whose underlying set is $\{(0,y): -1 \leq y \leq 1\}$. In this prime end, $(0,-1)$ is a major point, but any other point is a minor point.

A prime end which contains at least one accessible point and at least one non-accessible point is called a prime end of <u>second type</u>. Such is the prime end described above. It can be easily proved that the accessible point of a prime end E_g is a major point of E_g (see [14, p.353]). Hence, it follows that a prime end of second type is one which contains one accessible major point and at least one minor point.

A prime end of <u>third type</u> contains only non-accessible major points. Consider, for example, the region G (see Figure 29) bounded by the curves

$$x = 0, \ x = 1, \ y = \sin(\pi/x) + x$$

and

$$y = \sin(\pi/x) - x.$$

The prime end of G whose underlying set is the segment of the y-axis from $y = -1$ to $y = 1$ is of the third type.

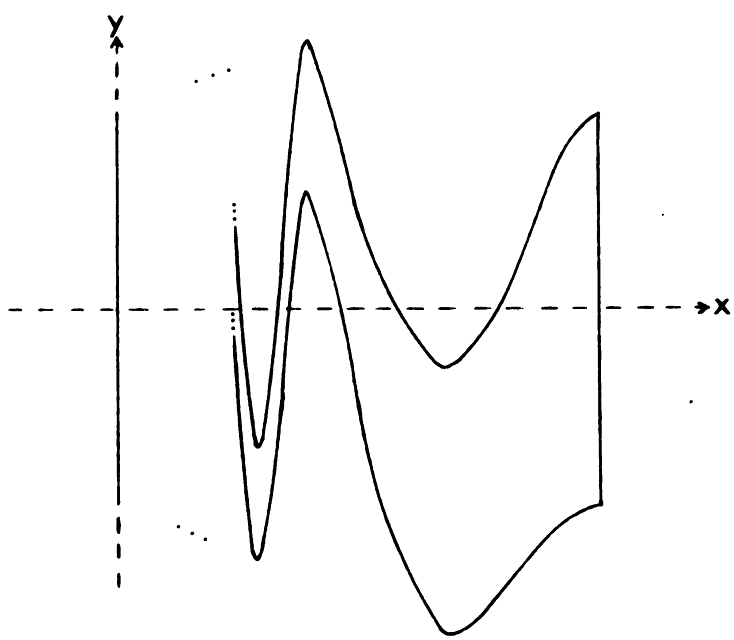


Figure 29

The fourth type are those prime ends which consist of non-accessible major points and some minor points as well. If we let G be the region bounded by the curves

$$x = 0, \ x = 1, \ y = \sin(\pi/x) + \frac{1}{2}$$

and

$$y = \sin(\pi/x) - \frac{1}{2},$$

then the prime end whose underlying set is the segment of the y-axis between -3/2 and 3/2 is of fourth type (see Figure 30). In this prime end, the entire segment of the y-axis between -1/2 and 1/2 consists of major points. To construct a chain of cross-cuts converging to any such point, say the point $(0,c)$, take those segments of the line $y = c$ which lie within the region G.

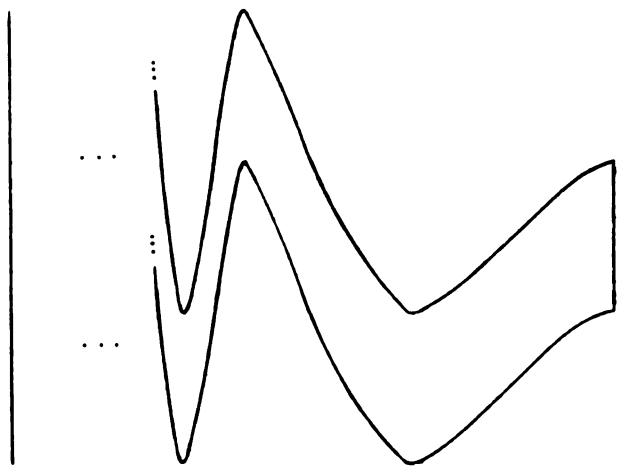
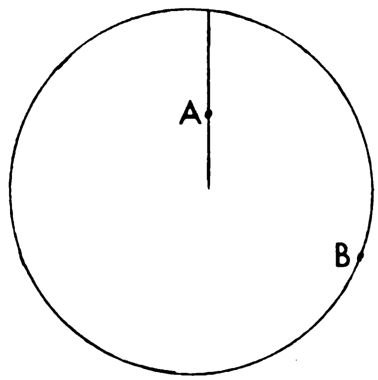


Figure 30

Carathéodory's goal was to characterize those regions whose boundary is the continuous, one-to-one image of a circle. Because of the close correspondence between the points of the boundary of a unit disk and the prime ends of a region, and because of the related convergence properties of each, the prime ends of first type play a central role in this characterization. However, to complete the characterization, the concept of multiplicity of a point is needed. A point of the boundary of a region G is a <u>simple point</u> of G if it is contained in only one prime end of G. Otherwise it is called a <u>multiple</u> (double, triple, etc.) <u>point</u> of E_g.

Consider the region whose boundary is the following set:



The point A is the underlying set of two prime ends of first type and is, therefore, a double point of each of these prime ends. The point. B, on the other hand, is a simple point of the prime end to which it belongs.

A characterization of simple points is given by the following theorem.

<u>Theorem 6.12</u>. A point A of the boundary γ of a region G is a simple point of γ iff every sequence of points that converges to A also converges to a prime end of G.

<u>Proof</u>: See [14, p.363].

The next result -- the main theorem of Carathéodory's paper -- shows the relationship of prime ends to Jordan curves.

<u>Theorem 6.13</u>. The boundary γ of a region G is homeomorphic to the unit circle iff γ contains only simple points and consists only of prime ends of first type.

The proof of this theorem shows that this result can be slightly generalized. Before discussing the generalization we present the proof essentially as Carathéodory gave it.

<u>Proof of Theorem 6.13</u>: The necessity of these conditions is easily seen.

To prove the sufficiency, suppose γ satisfies the two conditions. By Theorems 6.9 and 6.12, under any conformal mapping of G onto the unit disk $|z| \leq 1$, every sequence of points $P_1, P_2, \ldots$ of G which converges to a point A of γ corresponds to a sequence

$P_1', P_2', \ldots$ in $|z| < 1$ which converges to a point A' of the circle $|z| = 1$. Since every prime end of G consists of only a single point, it follows from the same two theorems cited above that to every sequence $Q_1', Q_2', \ldots$ in $|z| < 1$ converging to a point B' of $|z| = 1$ there corresponds a sequence $Q_1, Q_2, \ldots$ in G which converges to a well-defined point B of Υ. Thus the correspondence is one-to-one.

Let $A_1, A_2, \ldots$ be a sequence of points of Υ which converges to a point A, and let $A_1', A_2', \ldots$ and A' be their respective images under the conformal mapping of G onto the unit disk. For each value of n we can find a point P_n such that P_n and its image P_n' are within $1/n$ of A_n and A_n', respectively. The sequence $P_1, P_2, \ldots$ converges to A, therefore, $P_1', P_2', \ldots$ converges to A', and consequently $A_1', A_2', \ldots$ also converges to A'. Thus the mapping is continuous and the theorem is proved. □ [14, p.366]

The fact that the boundary points are simple was only used in proving that the mapping is one-to-one. Without that hypothesis the conformal mapping still gives a continuous correspondence between the points of the circle and those of the boundary G. Maria Torhorst, in her thesis (about 1920), pointed out this fact and restated Carathéodory's result as follows (see [80, p.63]):

Theorem 6.14. The boundary of a bounded simply connected region G is the continuous image of the closed unit interval iff all the prime ends of G are of first type.

Torhorst further proved:

Theorem 6.15. Let A be any point of the boundary γ of a simply connected region G. Then the following are equivalent:

(i) A is accessible from all sides with respect to G;

(ii) A is the accessible point of a prime end of G of first type;

(iii) γ is connected im kleinen at A.

Proof: See [80, p.63].

In short, Torhorst established the logical equivalence of all three then known characterizations. That is, the boundary γ of a bounded simply connected region G is a continuous image of the closed unit interval iff

(i) (Schoenflies): every point of γ is accessible from all sides with respect to G; or

(ii) (Carathéodory): all the prime ends of G are of first type; or

(iii) (Hahn-Mazurkiewicz): γ is connected im kleinen.

CHAPTER VII

EPILOGUE

The period between 1914 and 1920 saw very little development of the work of Hahn and Mazurkiewicz. The few results that we were able to locate were obtained by two closely associated American mathematicians R.L. Moore and J.R. Kline. (We are aware of the existence of another paper by Mazurkiewicz, written in 1916, but could not obtain a copy of it. The article was apparently written in Polish, anyway, and was not, as were many of Mazurkiewicz' other papers, translated into French or German.)

In 1917, Moore proved the following theorem.

Theorem 7.1. Every two points of a continuous curve are the endpoints of at least one simple arc lying entirely on that curve.

Proof: See [54, p.233].

The proof uses Hahn's definition of connectedness im kleinen as part of the characterization of continuous curve. This theorem is, of course, an extension of some of the earlier work of Mazurkiewicz and Janiszewski (see Theorem 4.4). Though Moore seems to have been unaware of the fact, Hahn had already proved a similar (if somewhat weaker) result in 1914 (see Lemma 5.8).

We have already mentioned (page 93) the next theorem -- one which Moore "rediscovered" in 1918.

Theorem 7.2. A bounded, simply connected, two-dimensional region has a simple closed curve as its boundary iff it is uniformly connected im kleinen.

Proof: See [55, p.364].

Denjoy had proved the sufficiency of the condition as early as 1911 (see Theorem 4.6), but Moore was the first to prove its necessity. He was evidently unaware of the works of Denjoy and Brouwer which we discussed in Chapter IV. Though he credits Hahn with the definition of uniform connectedness im kleinen, it was Moore himself who first formulated it as such and gave it its name.

In 1919, Moore added to Janiszewski's earlier work (see Chapter IV) the following result.

Theorem 7.3. Every bounded continuum that contains no continuum of condensation is connected im kleinen.

Proof: See [56, p.174].

The work of J.R. Kline on connectedness im kleinen prior to 1920 can be assumed up in the following entry from a 1918 issue of the Bulletin of the American Mathematical Society [39]:

"Dr. Kline proves the following theorem:

"A necessary and sufficient condition that a closed, connected, connected in kleinem [sic], plane point set should be a simple curve is that it divide its plane into two mutually exclusive domains."

Neither a proof nor an indication of the proof is included with the note. The paper to which the note refers was read at a meeting of the American Mathematical Society in New York, on April 27, 1918, but the text of that paper was apparently never published elsewhere.

From 1920 on, articles using connectedness im kleinen began to appear regularly in mathematical journals in Europe and the United States. The first volume of Fundamenta Mathematicae (1920) contains four such articles, including Mazurkiewicz' own summary and extension of his original papers which we have already discussed. The other three articles are a paper by Kuratowski [44], one by Sierpiński [79], and a joint paper by Janiszewski and Kuratowski [34].

In [44], Kuratowski gave a "purely topological" definition of "Jordan curve," that is, a definition using only the idea of limit point as the fundamental notion. To accomplish this, he characterized in topological terms the concept of a point of first genre as introduced by Mazurkiewicz. Before giving that characterization, we need the following preliminary definitions. A <u>semi-continuum</u> is any set S such that for any pair of points $x, y \in S$, there is a continuum $C \subset S$ containing x and y. If M is any set and p is a point of M, the <u>constituant of M determined by p</u> is the largest semi-continuum contained in M and containing p, that is, if C is the constituant of M determined by p and S is any other semi-continuum contained in M and containing p, then $S \subset C$. Let M be a non-empty subset of a continuum K. $p \in M$ is an <u>interior point of M relative to K</u> if it is not a limit point of $K - M$. M is a <u>domain relative to K</u> if every point of M is an interior point with respect to K. M is a <u>connected domain relative to K</u> if every point of M is an interior point with respect to K. M is a <u>connected domain relative to K</u> if it is a domain relative to K and each pair of points of M is contained in a subcontinuum of M.

<u>Theorem 7.4</u>. A point p of a continuum C is of first genre iff for every domain M (relative to C) containing p, p is an interior point (relative to C) of the constituant of M determined by p.

<u>Proof</u>: See [44, p.41].

As an immediate consequence of Theorems 7.4 and 5.4, Kuratowski arrives at the next result, which serves as his topological definition of a Jordan curve.

<u>Theorem 7.5</u>. A bounded continuum C is a Jordan curve iff every domain relative to C is the union of connected domains relative to C.

Out of 24 articles appearing in the first volume of Fundamenta Mathematicae, fourteen were the work of W. Sierpiński. The one we discuss here, [79], deals with an interesting property of point sets which is related to connectedness im kleinen. A set M (in E^n) is said to possess <u>property S</u> if for every positive number ϵ, every connected component of M can be expressed as a finite union of continua (in E^n) each having diameter $< \epsilon$.

In Sierpiński's formulation of the property, M was taken to be a continuum. The generalization we have given here is due to R.L. Moore, who, incidentally, was the first to call it property S (in honor of Sierpiński) (see [58]).

As the following theorem shows, property S is equivalent to connectedness im kleinen for bounded, closed, connected subsets of E^n.

Theorem 7.6. A continuum C is a Jordan curve iff it possesses property S.

Proof: See [79, p.44].

To see that the two concepts are not equivalent if the set is not closed, consider the half-open interval $T = (0,1]$. T is connected im kleinen but cannot be expressed as the union of a finite number of continua. R.L. Moore [58] has shown that property S is stronger than connectedness im kleinen but weaker than uniform connectedness im kleinen when applied to bounded subsets of E^n.

The fourth paper, [34], contains only a brief reference to connectedness im kleinen embodied in the following theorem on indecomposable continua.

Theorem 7.7. An indecomposable continuum is not connected im kleinen at any of its points.

Proof: See [34, p.217].

One of the most familiar results in the theory of connected im kleinen spaces is the following:

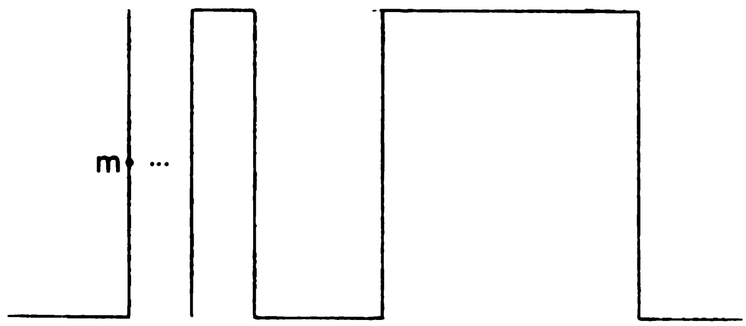
Theorem 7.8. The space R is connected im kleinen iff every component of an open subset of R is open.

Proof: See [19, p.113].

This result, first discovered and proved by Hahn in 1921 (see [23]), is also one of the earliest general results on connected im kleinen spaces in the literature. (Incidentally, this paper, too, appeared in one of the early issues of Fundamenta Mathematicae.) Hahn's original statement of the theorem was for metric spaces and his proof, using the idea of convergence of sequences, is metric dependent.

The stimulus for Theorem 7.8 can be found in the following examples. Let R be the space consisting of the real line with the closed segments $[\frac{1}{2n}, \frac{1}{2n-1}]$ $(n = 1,2,...)$ removed. R itself is an open set and all of the components of R are open (in R) except one -- the ray $(-\infty,0]$, since the point $x = 0$ is not an interior point of this ray. Of course, R is not connected. But consider the space R' composed of R (situated on the x-axis) together with the sets

$\{(0,y):0 < y \leq 1\}$, $\{(x,1):x \in [\frac{1}{2n},\frac{1}{2n-1}]$, for some $n = 1,2 \ldots\}$, and $\bigcup_{n=1}^{\infty} \{(1/n,y):0 < y < 1\}$:



R' is connected, but if m is the point $(0,\frac{1}{2})$ and $U = \{r \in R':d(r,m) < \frac{1}{2}\}$, then U is open in R' and the component of U along the y-axis is not open. Hahn observed that neither of these spaces is connected im kleinen at all of its points and was thus led to a formulation of Theorem 7.8.

As we view the origins of connectedness im kleinen in retrospect, we can see a definite pattern of development, one which has occurred quite often in the history of mathematics. An unexpected example challenged the traditional concept of "curve". The question was posed: "What is a curve?" Several attempts were made to answer that question -- most notable among them the definitions of Cantor and Jordan. In the process, more unusual examples were discovered and partial results were obtained to explain them. All of this occurred gradually and, out of the work done, there began to emerge a unifying theme -- indications of which at first appeared in isolated places here and there and in a rudimentary form. The underlying concept, connectedness im kleinen, was later clarified and put into focus by the independent efforts of two mathematicians -- Hahn and Mazurkiewicz.

We emphasize the fact that, as is true with the development of other mathematical ideas, connectedness im kleinen was the end product of a synthesizing process to which many different individuals contributed. Figure 31 illustrates some of the flow of ideas leading to, paralleling, and coming from the notion of connectedness im kleinen. On this chart, each solid arrow indicates a strong influence in the direction of the arrow, while each dashed arrow represents a significant, though indirect influence. Notice that there are two individuals at the side not tied in with the rest of the chart. This, too, is consistent with the pattern we have outlined. Denjoy and Nalli show no definite indication in their writings of any influence from the others represented in the diagram. However, as we pointed out earlier, they were working on the same kinds of problems that Carathéodory was researching -- namely, the character of the boundaries

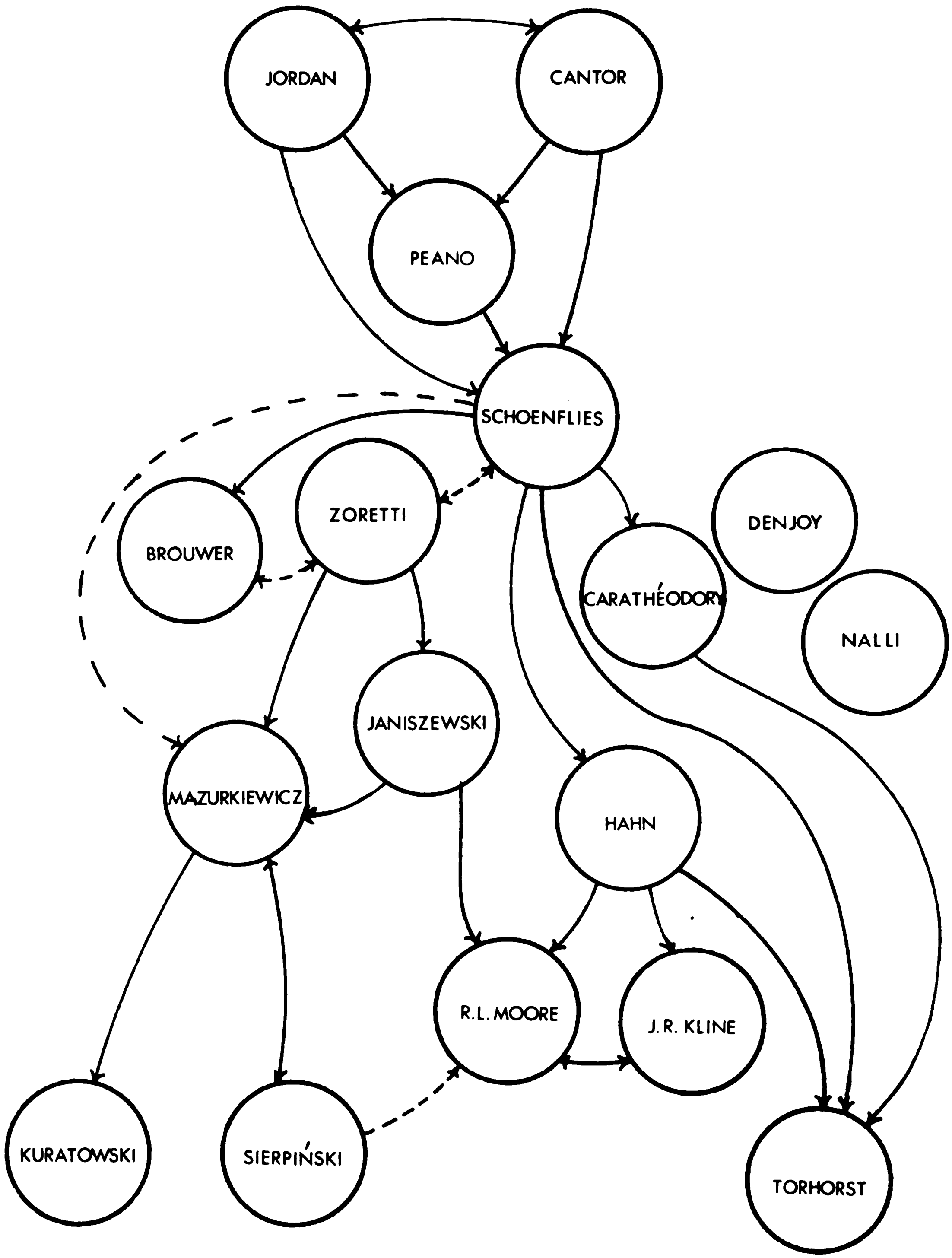

JORDAN
CANTOR
PEANO
SCHOENFLIES
ZORETTI
BROUWER
DENJOY
CARATHÉODORY
NALLI
JANISZEWSKI
MAZURKIEWICZ
HAHN
R.L. MOORE
J.R. KLINE
KURATOWSKI
SIERPIŃSKI
TORHORST

Figure 31

of simply connected domains -- and were motivated by the same questions.

Connectedness im kleinen, or its more modern version, local connectedness, has proved to be a useful mathematical concept through more than 60 years and is still an active tool of current research. It has passed the test of time. Perhaps in seeking out its origins we have gained an appreciation for the process by which mathematics itself is developed. This effort has certainly supported the observation of Wilder and Sarton that major mathematical concepts are rarely the complete invention of one man but are more often part of the "mathematical culture stream."

BIBLIOGRAPHY

BIBLIOGRAPHY

1. L. Bieberbach, Über den Jordanschen Kurvensatz, die Schoenfliesschen Sätze von Erreichbarkeit und Unbewalltheit und den Satz von der Invarianz des ebenen Gebietes, Deutsche Math. Verein. -- Jahresb., Abt. I, 22 (1913), 144-153.

2. C.B. Boyer, Historical Stages in the Definition of Curves, National Math. Magazine, 19 (1945), 294-310.

3. C.B. Boyer, A History of Mathematics, Wiley (1968).

4. L.E.J. Brouwer, Sur les continus irréductibles de M. Zoretti, Ann. Sci. de l'École Normale Supérieure, 27 (1910), 565-566.

5. L.E.J. Brouwer, Beweis des Jordanschen Kurvensatzes, Math. Ann., 69 (1910), 169-175.

6. L.E.J. Brouwer, Beweis der Invarianz des n-dimensionalen Gebiets, Math. Ann., 71 (1911), 305-313.

7. L.E.J. Brouwer, Beweis des Jordanschen Satzes für den n-dimensionalen Raum, Math. Ann., 71 (1911), 314-319.

8. L.E.J. Brouwer, Über Jordansche Mannifaltigkeiten, Math. Ann., 71 (1911), 320-327.

9. L.E.J. Brouwer, Beweis der Invarianz der geschlossenen Kurve, Math. Ann., 72 (1912), 422-425.

10. F. Cajori, A History of Mathematics, Macmillan (1919).

11. G. Cantor, Ein Beitrag zur Mannigfaltigkeitslehre, J. für reine und angewandte Math., 84 (1878), 242-258.

12. C. Carathéodory, Untersuchungen über die konformen Abbildungen, Math. Ann., 72 (1912), 107-144.

13. C. Carathéodory, Über die gegenseitige Beziehung der Ränder bei der konformen Abbildung des Inneren einer Jordanschen Kurve auf einen Kreis, Math. Ann., 72 (1912), 305-320.

14. C. Carathéodory, Über die Begrenzung einfach zusammenhängender Gebiete, Math. Ann., 73 (1913), 323-370.

15. E. Cesàro, Sur la représentation analytique des régions, et des courbes qui les remplissent, Bull. des Sci. Math., ser. 2, 21 (1897), 257-266.

16. A. Denjoy, Sur l'analysis situs du plan, Comptes Rendus de l'Acad. Sci. Paris, 153 (1911), 423-426.

17. A. Denjoy, Sur l'analysis situs du plan, Comptes Rendus de l'Acad. Sci. Paris, 153 (1911), 493-496.

18. A. Denjoy, Articles et mémoirs, Volume I, Gauthier-Villars (1955).

19. J. Dugundji, Topology, Allyn and Bacon (1966).

20. H. Hahn, Über die Abbildung einer Strecke auf ein Quadrat, Annali di Mat. Pura ed Appl., ser. 3, 21 (1913), 33-55.

21. H. Hahn, Über die allgemeinste ebene Punktmenge, die stetiges Bild einer Strecke ist, Deutsche Math. Verein. -- Jahresb., 23 (1914), 318-322.

22. H. Hahn, Mengentheoretische Characterisierung der stetigen Kurve, Akad. Wiss. Vienna -- Sitzungsb., 123 (1914), 2433-2489.

23. H. Hahn, Über die Komponenten offener Mengen, Fund. Math., 2 (1921), 189-192.

24. H. Hahn, Über die stetigen Kurven der Ebene, Math. Zeit., 9 (1921), 66-73.

25. H. Hahn, Über irreduzible Kontinua, Akad. Wiss. Vienna -- Sitzungsb., 130 (1921), 217-250.

26. D.W. Hall, and G.L. Spencer II, Elementary Topology, Wiley (1955).

27. F. Hausdorff, Grundzüge der Mengenlehre, Chelsea (1914).

28. D. Hilbert, Ueber die stetige Abbildung einer Linie auf ein Flachenstück, Math. Ann., 38 (1891), 459-460.

29. J.G. Hocking, and G.S. Young, Topology, Addison-Wesley (1961).

30. Z. Janiszewski, Contribution à la géométrie des courbes planes générales, Comptes Rendus de l'Acad. Sci. Paris, 150 (1910), 606-609.

31. Z. Janiszewski, Sur la géométrie de lignes cantoriennes, Comptes Rendus de l'Acad. Sci. Paris, 151 (1910), 198-201.

32. Z. Janiszewski, Sur les continus irréductibles entre deux points, J. de l'École Polytechnique, ser. 2, 16 (1912), 76-170.

33. Z. Janiszewski, Über die Begriffe "Linie" und "Fläche", Proc. Fifth Int. Cong. of Mathematicians, Volume 2, (1913), 126-128.

34. Z. Janiszewski, and K. Kuratowski, Sur les continus indécomposables, Fund. Math., 1 (1920), 210-222.

35. F.L. Jones, A History and Development of Indecomposable Continua Theory, Ph.D. Thesis, Michigan State University (1971).

36. C. Jordan, Cours d'analyse de l'École Polytechnique, Volume I, 2nd ed., Gauthier-Villars (1893).

37. H.C. Kennedy, Selected Works of Giuseppe Peano, Univ. of Toronto Press (1973).

38. J.R. Kline, Concerning Approachability of Simple Closed and Open Curves, Trans. Am. Math. Soc., 21 (1920), 451-458.

39. J.R. Kline, Note 13, Bull. Am. Math. Soc., 24 (1918), 471.

40. J.R. Kline, A Theorem Concerning Connected Point Sets, Fund. Math., 3 (1922), 238-239.

41. J.R. Kline, What Is the Jordan Curve Theorem?, Am. Math. Monthly, 49 (1942), 281-286.

42. K. Knopp, Einheitliche Erzeugung und Darstellung der Kurven von Peano, Osgood und v. Koch, Archiv der Math. and Phys., ser. 3, 26 (1917), 103-115.

43. H.v. Koch, Une méthode géométrique élémentaire pour l'étude de certaines questions de la théorie des courbes planes, Acta Math., 30 (1905), 145-174.

44. K. Kuratowski, Une définition topologique de la ligne de Jordan, Fund. Math., 1 (1920), 40-43.

45. K. Kuratowski, Topology, Volume I, Academic Press (1966).

46. K. Kuratowski, Topology, Volume II, Academic Press (1968).

47. H. Lebesgue, Sur le problème des aires, Bull. Soc. Math. de France, 31 (1903), 197-203.

48. S. Mazurkiewicz, Sur la théorie des ensembles, Comptes Rendus de l'Acad. Sci. Paris, 151 (1910), 296-298.

49. S. Mazurkiewicz, O arytmetyzacyi continuów, Comptes Rendus de la Soc. Sci. Varsovie, Classe III, 6 (1913), 305-311. (French translation available in [52].)

50. S. Mazurkiewicz, O arytmetyzacyi continuów II, Comptes Rendus de la Soc. Sci. Varsovie, Classe III, 6 (1913), 941-945. (French translation available in [52].)

51. S. Mazurkiewicz, Sur les lignes de Jordan, Fund. Math., 1 (1920), 166-209.

52. S. Mazurkiewicz, Travaux de topologie et ses applications, l'Acad. Polonaise des Sciences (1969).

53. E.H. Moore, On Certain Crinkly Curves, Trans. Am. Math. Soc., 1 (1900), 72-90.

54. R.L. Moore, A Theorem Concerning Continuous Curves, Bull. Am. Math. Soc., 23 (1917), 233-236.

55. R.L. Moore, A Characterization of Jordan Regions by Properties Having No Reference to Their Boundaries, Proc. Nat. Acad. Sci., 4 (1918), 364-370.

56. R.L. Moore, Continuous Sets That Have No Continuous Sets of Condensation, Bull. Am. Math. Soc., 25 (1919), 174-176.

57. R.L. Moore, On the Relation of a Continuous Curve to Its Complementary Domains in Space of Three Dimensions, Proc. Nat. Acad. Sci., 8 (1922), 33-38.

58. R.L. Moore, Concerning Connectedness im Kleinen and a Related Property, Fund. Math., 3 (1922), 232-237.

59. R.L. Moore, Report on Continuous Curves from the Viewpoint of Analysis Situs, Bull. Am. Math. Soc., 29 (1923), 289-302.

60. P. Nalli, Sulla definizione di dominio, Suppl. Rend. del Circ. Mat. di Palermo, 6 (1911), 31-32.

61. P. Nalli, Sopra una definizione di dominio piano limitato da una curva continua, senze punti multipli, Rend. del Circ. Mat. di Palermo, 32 (1911), 391-401.

62. E. Netto, Beitrag zur Mannigfaltigkeitslehre, J. für reine und angewandte Math., 86 (1879), 263-268.

63. W.F. Osgood, Allgemeine Theorie der analytischen Funktionen, Enc. der Math. Wiss., II B, (1901), 1-114.

64. W.F. Osgood, A Jordan Curve of Positive Area, Trans. Am. Math. Soc., 4 (1903), 107-112.

65. P. Painlevé, Sur la théorie de la représentation conforme, Comptes Rendus de l'Acad. Sci. Paris, 112 (1891), 653-657.

66. G. Peano, Sur une courbe, qui remplit toute une aire plane, Math. Ann., 36 (1890), 157-160.

67. G. Polya, Über eine Peanosche Kurve, Bull. Int. de l'Acad. Sci. Cracovie, ser. A, Sciences Mathématiques, (1913), 305-313.

68. A. Rosenthal, Über Peanoflächen und ihren Rand, Math. Zeit., 10 (1921), 102-104.

69. H.L. Royden, Real Analysis, Macmillan (1968).

70. G. Sarton, The Study of the History of Science, Dover (1957).

71. A.M. Schoenflies, Die Entwicklung der Lehre von den Punktmannigfaltigkeiten, Part I, Deutsche Math. Verein -- Jahresb., 8_2 (1900), 1-250.

72. A.M. Schoenflies, Über einen grundlegenden Satz der Analysis Situs, Akad. der Wiss., Göttingen, Math.-Phys. Klasse, Nachr. (1902), 185-192.

73. A.M. Schoenflies, Über den allgemeinsten Begriff der ebenen stetigen Kurve, Akad. der Wiss., Göttingen, Math.-Phys. Klasse, Nachr. (1907), 28-49.

74. A.M. Schoenflies, Über den allgemeinsten Begriff der ebenen stetigen Kurve, Akad. der Wiss., Göttingen, Math.-Phys. Klasse, Nachr. (1907), 299-320.

75. A.M. Schoenflies, Die Entwicklung der Lehre von den Punktmannigfaltigkeiten, Part II, Deutsche Math. Verein. -- Jahresb., Ergänzungsbände, 2 (1908), 1-331.

76. H.A. Schwartz, Zur Integration der partiellen Differentialgleichung $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$, J. für reine und angewandte Math., 74 (1872), 218-253.

77. W. Sierpiński, Sur une nouvelle courbe continue qui remplit toute une aire plane, Bull. Int. de l'Acad. Sci. Cracovie, ser A, Sciences Mathématiques, (1912), 462-478.

78. W. Sierpiński, Sur une courbe non quarrable, Bull. Int. de l'Acad. Sci. Cracovie, ser A, Sciences Mathématiques, (1913), 254-263.

79. W. Sierpiński, Sur une condition pour qu'un continu soit une courbe jordanienne, Fund. Math., 1 (1920), 44-60.

80. M. Torhorst, Über den Rand der einfach zusammenhängenden ebenen Gebiete, Math. Zeit., 9 (1921), 44-65.

81. O. Veblen, Theory on Plane Curves in Non-Metrical Analysis Situs, Trans. Am. Math. Soc., 6 (1905), 83-98.

82. R.L. Wilder, Topology of Manifolds, Am. Math. Soc. Coll. Pub. No. 32 (1949).

83. R.L. Wilder, The Origin and Growth of Mathematical Concepts, Bull. Am. Math. Soc., 59 (1953), 423-448.

84. K. Yoneyama, Theory of Continuous Sets of Points, Tôhoku Math. J., 12 (1917), 43-158.

85. G.C. Young, On the Form of a Certain Jordan Curve, Quarterly J. of Pure and Applied Math., 37 (1906), 87-91.

86. W.H. Young and G.C. Young, The Theory of Sets of Points, Chelsea (1972).

87. L. Zoretti, Un Théorème de la théorie des ensembles, Bull. Soc. Math. France, 37 (1909), 116-119.

88. L. Zoretti, La notion de ligne, Ann. Sci. de l'École Normale Supérieure, ser 3, 26 (1909), 485-497.

89. L. Zoretti, Sur les propriétés des lignes cantoriennes, Comptes Renus de l'Acad. Sci. Paris, 150 (1910), 1505-1506.

90. L. Zoretti, Sur la notion de ligne, Comptes Rendus de l'Acad. Sci. Paris, 151 (1910), 201-203.

91. L. Zoretti, Remarque au sujet de la remarque de M. Brouwer, Ann. Sci. de l'École Normale Supérieure, ser. 3, 27 (1910), 569.

92. L. Zoretti, Sur la représentation analytique d'un continu irréductible, Bull. Soc. Math. France, 39 (1911), 246-250.

93. L. Zoretti, Les ensembles de points, Enc. des Sci. Math., II_1 (1912), 113-170.

94. L. Zoretti, Contribution à l'étude des lignes cantoriennes, Acta Math., 36 (1912), 241-268.

95. L. Zoretti, and H. von Mangoldt, Les notions de ligne et de surface, Enc. des Sci. Math., III_1 (1911), 152-184.