FINDING THE PATH

THEMES AND METHODS FOR THE TEACHING OF MATHEMATICS IN A WALDORF SCHOOL

Bengt Ulin

Published by The Association of Waldorf Schools of North America

Swedish Text: ATT FINNA ETT SPÅR

Motiv och metoder i matematikundervisningen

Edited by David Mitchell

Library of Congress Cataloging-in-Publication Data

Ulin, Bengt

FINDING THE PATH

THEMES AND METHODS FOR THE TEACHING OF MATHEMATICS IN A WALDORF SCHOOL

Translation of: *Att Finna Ett Spår.*

ISBN: 0-9623978-1-4

Printed in the United States of America.

Designed by Robin Brooks

ACKNOWLEDGEMENT

The editor wishes to express gratitude to Bengt Ulin, a mathematics teacher at the Kristoffer School in Stockholm, Sweden, for making it possible for this edition of his book, *Att Finna ett Spår*, to be available to English speaking audiences. I would also like to thank members of the A.W.S.N.A. Publications Committee for their assistance, most notably to Susan Demanett for proofreading and to Jan Baudendistel for distribution.

February, 1991
Wilton, N.H.

David Mitchell

CONTENTS

PREFACE TO THE THIRD SWEDISH EDITION

For the individual who wishes to find out what mathematics is all about there are no simple shortcuts. One must work with the mathematics itself. The job can be made easier if one succeeds in finding a cicerone with feeling for both the subject and the innately human, a guide who can stimulate one's thoughts and joy of discovery.

Bengt Ulin is such a cicerone. He has understood that mathematical discovery need not at all be seen as something for only a small exclusive group of great mathematicians, whose results in polished and suitably humble form are presented to students as "facts." He understands that within mathematics there is something for each of us to discover. There is then, of course, nothing new for humanity. There is, in a sense, something more important than that, something which gives the student true knowledge, which shows how knowledge takes form within the student, how thinking develops, and what creativity means.

Just as in the case of the person who wishes to know what travel is all about, being presented a few facts is not enough. One must tread the road himself. A good travel guide is usually a help. *Finding the Path* is an excellent guide.

It ought to find a very broad readership. Within the school system it will be useful for both active teachers and new teachers-to-be, at different levels, as well as for many high school students. Here I am thinking not primarily of those who go the finals in mathematical contests but rather of a large group who might become interested in mathematics if they only got the right stimulus. Apart from the school system I believe there are many of the general public who through this book could find a path, perhaps even a whole new hobby, in mathematics.

Bengt Ulin follows a thinking tradition which leads one to associations with George Pólya. He encourages experimentation while not forgetting precision when it is time for it. Whether this is then called Waldorf pedagogy or something else is of no matter. Of importance is that we in this book have examples showing that mathematics does not need to be a mechanical manipulation of symbols according to certain rules which must be memorized, that mathematics can lead to other than a fixation with answers, a phenomenon which hinders the student from seeing the actual road. *Finding the Path* helps the reader to discover patterns and structure and leads the reader to much of the beauty which mathematics has to offer.

It is pleasing to note that through this new edition *Finding the Path* will become available to a larger public, and I can only wish all readers good luck on their fascinating journey under the enthusiastic leadership of the very knowledgeable Bengt Ulin.

Andrejs Dunkels
Professor of Mathematics
University of Luleå

1

INTRODUCTION

The presentations in this book are built on experiences from mathematics teaching in the grades 7-12 at the Kristoffer School, a Rudolf Steiner School, and from teaching at the Rudolf Steiner Seminar at Jaerna (Sweden).

Visitors of the school, of the teacher seminar, and of Waldorf school exhibitions have asked many questions about the teaching in the Waldorf school and about mathematics. Some visitors have desired printed material.

During the evaluation of the Kristoffer School by the Swedish state school authorities 1976/77, the question about publication of the methods in some subjects was actualized, especially since the evaluation could not comprise mathematics and science to an extent which had been proper.

Largely as a result of the stimulating interest of Karl-Georg Ahlströem, professor at the Department of Pedagogy at the University of Uppsala and leader of the evaluation group, I began to collect glimpses and experiences from the lessons.

Dr. Georg Unger, leader of the Institute for Mathematics and Physics at Dornach, Switzerland, gave me important impulses. Ingemar Wik (University of Umeå), Hans Brolin (The High school for teachers, Uppsala), and Sven-Erik Gode (the publishing firm Natur och Kultur, Stockholm) took the trouble to read the first version, and I am thankful for their advice concerning the work which was to follow.

The book is primarily written for younger mathematics teachers or teachers-to-be. It does not offer a pedagogical collection of recipes, but should simply give examples of how one might engage the pupils in heterogeneous (undifferentiated) classes.

Most of the book is taken up by Chapter 3, which shows a number of themes from the teaching. In the Waldorf school, mathematics is taught in all grades, during periods, every morning for some weeks during each period, and during fixed, weekly exercise hours.

What is presented in the twelve sections of Chapter 3 is not reports from the whole period or a whole lesson but examples of themes able to engage the pupils.

With the purpose of also enabling generally interested readers with varying prerequisites to follow the expositions, I have simplified the text as much as possible and avoided unnecessary mathematical terms. Hoping to reach different categories of readers, I have given Chapter 3 a lot of space. Its sections also offer pedagogical material for the remarks in Chapters 4–8.

The presentation is as little a textbook in mathematics as a pedagogical handbook. There are several domains of school mathematics which do not appear. The purpose has not been to give a total view of the school curriculum but to contribute to the pedagogical development. I am aware that experienced teachers will recognize much of the content, but there may be variations of some themes which might be interesting even for them, perhaps aspects on teaching in rather heterogeneous classes.

Besides the persons already mentioned, I want to thank my colleagues Arne Nicolaisen (Oslo), Lars Hallqvist and Rüdiger Neuschütz (Bromma) for advice and control work.

At last I wish to thank the Swedish state school authorities for financing the first edition as a project work, entitled "Methods in Mathematics Teaching from the View of Waldorf Pedagogy."

2

MATHEMATICS AS A PATH FOR DEVELOPMENT OF THINKING — PAST AND PRESENT

2.1 Foreign Cultures

The papyrus scrolls with mathematics texts which were found in Egypt bear witness to well developed abilities in problem-solving among an elite in ancient Egypt. The texts, particularly the Rhind Papyrus and the Moscow Papyrus, date back to originals from the Middle Kingdom (2000-1800 B.C.) and thus are almost 4000 years old. The Rhind Papyrus begins by explaining that its contents concern the art of "penetrating into things" and that it will provide "knowledge of everything existing, of all secrets." In actual fact the Rhind Papyrus is a collection of recipes for calculating grain requirements, distributing wages, farm field areas and storehouse volumes, conversion of units, etc. In short, the text is a collection of methods for the solution of various practical problems.

B.L. van der Waerden believes these mathematical texts were intended for teaching in a school for scribes, the royal public servants who were the master calculators and "undersecretaries" to the Pharaoh. A limestone sculpture from the 5th dynasty of the Ancient Kingdom (2500 B.C.), now in the Louvre, shows the intensity of concentration of a scribe as he sits ready to take notes. The scribe could lose his life if he made an error.

The ancient Egyptians were particularly successful in geometry. They used the very good approximation

$$4 \cdot \left(\frac{8}{9}\right)^2 = 3.16049...$$

for π and they performed the masterly feat of calculating the volume of a truncated square pyramid using the correct formula

$$V = (a^2 + ab + b^2) \cdot \frac{h}{3}$$

where a and b are the sides of the square top and bottom sections and h is the height between them.

The Babylonians developed algebra surprisingly far. Cuneiform texts reveal that they could solve quadratic equations and knew of methods for treating certain third degree equations. They solved systems of equations, even of the second degree, and succeeded in developing a number of arithmetic formulas, including among others the formula for sums of squares (see Section 3.5.6).

In geometry the Babylonians were well acquainted with the concept of similarity and with triplets of numbers which form right triangles (Pythagorean numbers). They even calculated the volume of the prism and the cylinder.

According to research in the history of mathematics the Greeks had probably acquainted themselves with the mathematics of both Egyptians and Babylonians. Much points to Greeks having spent considerable time in Egypt and in the Tigris-Euphrates valley and there having the opportunity of thoroughly studying that which had been developed. Various accounts report of Thales, Pythagoras, Democritus and Eudoxus — all prominent mathematicians in ancient Greece — undertaking travels to Egypt and Babylonia.

This in no way, however, entitles the conclusion that Greek mathematics was simply a product of what they had found in other cultures. Van der Waerden quotes Plato in the posthumously published dialogue "Epinomis", words which van der Waerden finds particularly apt:

> ...whatever Greeks acquire from foreigners is finally turned by them into something nobler.

What the Greeks could make use of in the area of mathematics were methods for problem-solving, formulas, and reference tables — in short, instructions and data. But how had the Egyptians and Babylonians come to these results? Were they reliable? Among the ancient works were incorrect formulae and methods. For example, the Babylonians calculated the volume of a truncated cone with the formula:

$$V = \frac{3h}{2}(R^2 + r^2) \cdot$$

It was the Greeks' great service that they took the step from calculation to mathematics. They sought proofs for all of the results which they had come upon and developed the art of problem-solving to an eminent degree. They showed an impressive ability at finding constructive methods and at developing different forms of proof when needed to give a rigorous foundation for constructions or other methods of calculation. Definitions, assumptions, and axioms were formulated — all with an admirable precision.

It was Thales (ca: 600 B.C.) who brought proof into geometry. That which he received from the East and the South were results which had once been borne up by a living culture but which in his time existed only as document. From these collections of formulae, Thales structures a logical geometry. Consciousness makes its entry into mathematics: "one knows that one knows."

From studies of the pyramids we know that the Egyptians were good at applying geometry, as early as 2000 B.C.; they could build geometrical constructions on even the enormous scale required for the pyramids. Nor were th Greeks one-sided theoreticians. The 1 kilometer long Eupalinos tunnel through the Kastro mountain on Samos, approximately 530 B.C., containing a water channel arranged with ventilation shafts, speaks for itself. It was dug from both directions simultaneously! The two work crews met in the middle of the mountain with an error of less than 10 meters sideways and 3 meters vertically.

For the Greeks such practical tasks were, however, a by-product. The important thing for them was to develop thinking. The Pythagorean School, like Thales, attached great importance to the development of logical structure in geometry and proved with the aid of parallel lines, among other things, that the sum of angles in a triangle is 180°. They were well acquainted with the right triangle and developed a system of representing problems in arithmetic as geometrical problems — the reverse of Descartes' idea using a co-ordinate system to transform geometrical problems into algebraic ones.

Of major importance to the Greeks were clarity and lucidity. They were fond of summarizing a geometrical proof with a figure and the simple text: "Observe!" But for the Pythagoreans it was also a logical necessity to transform arithmetical problems into geometrical ones: they had not mastered the irrational numbers. The Pythagoreans' whole system of thought was long based on the belief that all numbers are either whole or made up of ratios of whole numbers (we call such numbers "rational"). When the study of the square's diagonal and the proportions in the regular pentagram led to the discovery of other "non-expressible" numbers, Pythagorean mathematics came to a crisis. It turned out that in geometry, with the aid of compass and straight edge, one could very easily construct lengths corresponding to certain irrational numbers; for example, the diagonal of a square in the case of $\sqrt{2}$.

With this, Greek mathematics had come into deep water, where neither inherited knowledge nor studies of the physical world were of any help. Mathematics came to be the science, where thought's own powers were developed, independent of sensory knowledge. Plato (427-348 B.C.) writes in his famous work "Republic":

> Through mathematics is the instrument of the soul cleansed as though awakened to a new life force in a tempering fire; while other occupations consume it and remove it from its power of sight, a power which would, however, be far more deserving of retention than a thousand bodily eyes, since only through such a mental instrument may the truth be seen.

Plato had been initiated into Pythagorean mathematics and other exact sciences by Archytas from Taras, the mathematician who solved the famous doubling of the cube problem from Delos through an ingenious spatial construction.

According to a number of Greek sources, the inhabitants of Delos had been advised by an oracle to double the size of an altar in order to rid themselves of a pestilence which prevailed upon their island. The architects were said to have come to despair to Plato, who told them that the oracle meant to criticize the Greeks for neglect of mathematics and geometry. For the mathematicians the problem became finding the side of a cube which has twice the volume of a given cube. Two other classical problems come from Greece: to construct a square with the same area as a given circle ("squaring the circle") and to divide a given angle into three equal parts ("trisecting an angle").

These three problems came to play an exceptional role not only in Greek culture but also in the continuing development of mathematics. Not until the 1800's was it proven that neither the cube's doubling, nor the circle's squaring, can be carried out using *only* compass and straight

edge, and that trisection of an angle with these instruments is only possible for certain special angles but not in general. The problem of squaring the circle was solved first in 1882 when Ferdinand Lindemann was able to show that $\sqrt{\pi}$ is a transcendental number.

Greek mathematicians succeeded in solving the three classical problems with *other* tools than only ruler and compass. In the previously mentioned spatial construction for doubling the cube, Archytas implicitly makes use of the principle of continuity. This and other solutions to these problems are recounted by van der Waerden; for example, the use of ready-made curves. Let us here examine a few examples of solutions of a "mechanical" nature. Figure 2.1.1 shows two lines at right angles to each other with intersection at O. A and B are two given points located such that $OB = 2 \cdot OA$.

If we set OA = 1 unit, then OB = 2 units. We wish to find the side X of a cube which has a volume of 2 units (twice as large as a given cube with volume 1 unit) and seek, therefore, to construct the length $X = \sqrt[3]{2}$.

If we can construct points X and Y in Figure 2.1.2 so that AX and BY are at right angles to XY, then it follows from similar triangles that x (= OX) and y (=OY) satisfy the relations

$$\frac{1}{x} = \frac{x}{y} = \frac{y}{2}$$

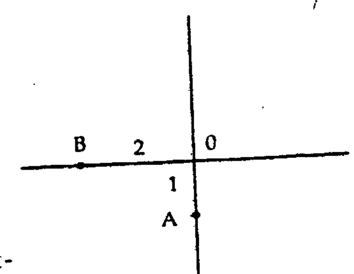


Figure 2.1.1

from which it follows that

$$y = x^2 \text{ and } xy = 2.$$

Consequently we have $x^3 = 2$, i.e. $x = \sqrt[3]{2}$.

OX thus gives the side of the sought-after cube. A Greek showed how one can "mechanically" construct X: one makes use of two U-shaped right angles as in Figure 2.1.3, of which the one slides inside the other. The larger right angle is placed so that the inner edge passes through B and the corner H falls on the y-axis OY. Apart from this it may be freely rotated. At the same time the smaller U-angle is moved so that its outer edge k goes through A and the corner K falls on the x-axis. When these two

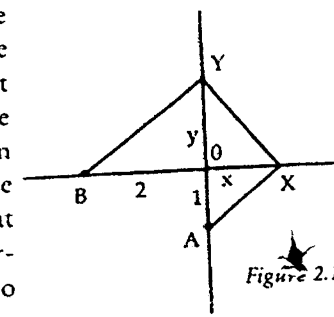


Figure 2.1.2

conditions are satisfied simultaneously, k's intersection with the x-axis will give the point sought, X.

In a dramatically written dialogue, "Platonicus," Eratosthenes has Plato utter the following criticism of Archytas, Eudoxus, and Menaichmos:

> Have you found mechanical solutions? There's no trick in that; even I who am no mathematician can do that. — With such methods the good in geometry will fall to ruin, and the eye will be directed from the purely geometrical to merely observable things...

For Plato it was obviously important that solutions be purely logical — constructive and preferably carried out within constructions which may be done exactly with straight edge and compass. These tools were merely aids with which the solution might be made observable to the senses. The important thing was the train of thought, which showed that the solution had been found and was correct. Let us now have a look at a problem which admits a solution with compass and straight edge and which in a typical manner satisfies Plato's demand for freedom from worldly tools:

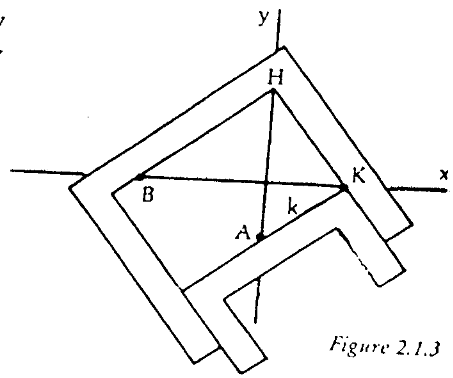


Figure 2.1.3

A circle C and a point P (not the center) are given. How can one draw a chord through P such that the chord has a given length k? (Figure 2.1.4)

What do we need to construct? We need to find an end-point of the desired chord. Since the chord must go through P, we have then only to lay the straight edge through P and the end-point and draw the line.

How shall we find an end-point? We have it, if we can determine the length of one of the two subdivisions of the chord lying on either side of P. How can we get such a length?

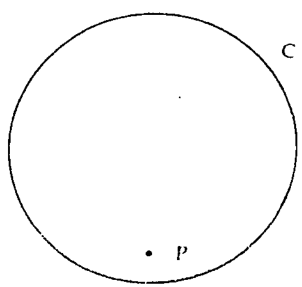


Figure 2.1.4

We utilize now the circle's symmetry of rotation and draw a chord k′ with length k anywhere on the circle: using the compass we mark off the distance k starting from an arbitrary point X and thereby get the other end-point Y on the circle (Figure 2.1.5). Can we rotate this chord, so that it passes through P? Or can we rotate P to a corresponding point P′?

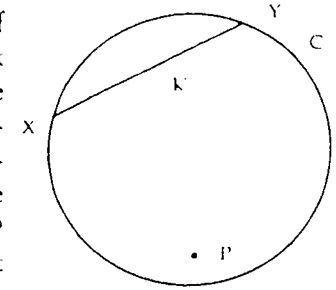


Figure 2.1.5

The latter is easily done with the compass placed at circle center and P, drawing the concentric circular arc around until it intersects the chord k′. We call the intersection P′(Figure 2.1.6).

We may now measure one of the two chord sublengths, say the shortest one, on k′. The length we get there can then be marked off from P out onto the circle and we have found an end-point of the desired chord. (The length may be marked off in two ways on the circle, the one solution is a mirror image of the other.)

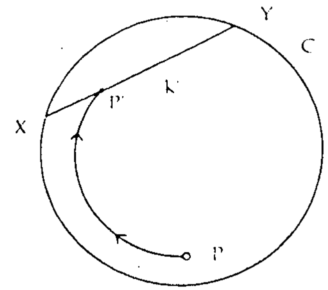


Figure 2.1.6

The discussion we have just carried out utilized the figures only to make the train of thought more easily understandable for another person. And the straight edge and compass have only served in the practical realization of the construction. The whole line of reasoning is conveyed in the world of thought.

It was such pure mathematics which Plato wanted to encourage. Mathematics would be a training ground for pure thinking, which need not rely on sensory perception (as it does in the case of "mechanical" solutions). This strictness in Greek geometry has been one of the major foundations for that logical clarity, that exactitude, and that consciousness which has been built up in western science, especially in the last 400 years.

Let us look at another example, taken from Archimedes (287?-212 B.C.), who lived considerably later than Plato and who must be seen as one of the greatest mathematicians of all time. Archimedes shows how a given angle can be trisected using a mechanical method.

In Figure 2.1.7 the given angle v is placed at the centre of a circle C of arbitrary radius. Archimedes extends the one angle leg MA (the line a) outside the circle and seeks a point X on that line so that the portion of the line segment BX lying outside the circle is of the same length as the radius. We can mark off the radius' length on the ruler and then mechanically adjust the ruler's position through B and to the line a so that the condition is satisfied (Figure 2.1.8). Then the angle u between BX and the line a will be one-third of this angle v.

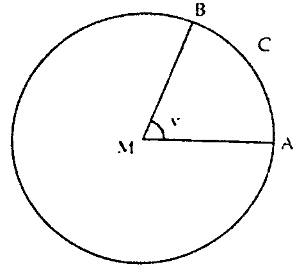


Figure 2.1.7

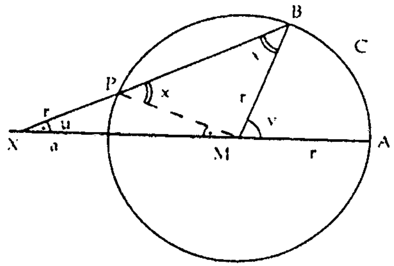


Figure 2.1.8

It is not difficult to prove this with the theorem on external triangles: using the nomenclature in the figure we have

$$x = 2u \quad \text{and} \quad v = x + u$$

from which $v = 3u$.

Archimedes was not the one to take lightly the logical, proof-finding side of mathematics, but he had a surprisingly ingenious talent for getting ideas and heuristic methods of problem solution. It is interesting to see how often he is led to pictures from mechanic, when he works with a problem. Archimedes describes, in his book *The Method*, how he approached certain difficult problems such as the area of a parabolic segment, the volume of a sphere, and the center of gravity of a hemisphere.

> "Certain things," writes Archimedes, "first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards, because their investigation by the said method did not furnish an actual demonstration. But it is, of course, easier when we have previously acquired, by the method, some knowledge of the question, to supply the proof than it is to find it without any previous knowledge."

Archimedes gave here an approach which would show itself to be very fertile. Most of the major results in mathematics have been achieved with an approach which is similar, in principle, to that given by Archimedes, even if the intuitive means of finding solutions have taken many expressions other than analogies in mechanics.

Archimedes' description also points out something important about mathematical activity: proof is the concluding phase. We will return to this in various contexts later on but for the moment simply ask: in what ways can we today make use of mathematics to develop thinking during school years?

2.2 Our Times

Technical development during recent decades places important questions before the "common man." Ignoring the complex of problems around war, the arms race, weapons, etc., we might name questions concerning the world's food production, its energy requirements and the necessity of environmental policy, particularly to safeguard the earth's own resources. Other important problems might be brought forth. Not least with energy and environmental questions have the difficulties in relations between experts and laymen become highly current. Politicians must here often be regarded as laymen. For him who is neither expert nor politician, but has given some thought to the importance of a particular problem, there are two extremes to take: one relies entirely on the experts and politicians, or one troubles oneself to penetrate into the available facts, studies, and evaluations. Depending on abilities, some laymen may go even further and more or less become experts.

Should we leave the thinking to the experts? Would not that be a form of resignation, authority worship, or the easy way out — all rather antiquated in our day?

The need in each of us to be able to think critically has certainly never been so current as it is today. Many scandals such as the neurosedyne catastrophe in Sweden might have been avoided if awareness had been greater of the role of one's own thinking. The individuals entrusted

In Figure 2.1.7 the given angle v is placed at the centre of a circle C of arbitrary radius. Archimedes extends the one angle leg MA (the line a) outside the circle and seeks a point X on that line so that the portion of the line segment BX lying outside the circle is of the same length as the radius. We can mark off the radius' length on the ruler and then mechanically adjust the ruler's position through B and to the line a so that the condition is satisfied (Figure 2.1.8). Then the angle u between BX and the line a will be one-third of this angle v.

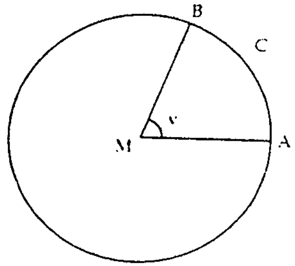


Figure 2.1.7

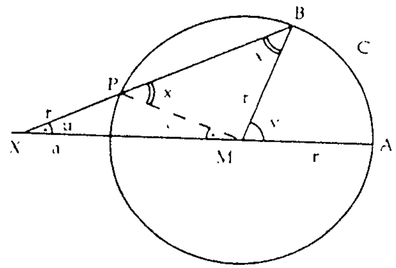


Figure 2.1.8

It is not difficult to prove this with the theorem on external triangles: using the nomenclature in the figure we have

$$x = 2u \quad \text{and} \quad v = x + u$$

from which v = 3u.

Archimedes was not the one to take lightly the logical, proof-finding side of mathematics, but he had a surprisingly ingenious talent for getting ideas and heuristic methods of problem solution. It is interesting to see how often he is led to pictures from mechanic, when he works with a problem. Archimedes describes, in his book *The Method*, how he approached certain difficult problems such as the area of a parabolic segment, the volume of a sphere, and the center of gravity of a hemisphere.

> "Certain things," writes Archimedes, "first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards, because their investigation by the said method did not furnish an actual demonstration. But it is, of course, easier when we have previously acquired, by the method, some knowledge of the question, to supply the proof than it is to find it without any previous knowledge."

Archimedes gave here an approach which would show itself to be very fertile. Most of the major results in mathematics have been achieved with an approach which is similar, in principle, to that given by Archimedes, even if the intuitive means of finding solutions have taken many expressions other than analogies in mechanics.

Archimedes' description also points out something important about mathematical activity: proof is the concluding phase. We will return to this in various contexts later on but for the moment simply ask: in what ways can we today make use of mathematics to develop thinking during school years?

2.2 Our Times

Technical development during recent decades places important questions before the "common man." Ignoring the complex of problems around war, the arms race, weapons, etc., we might name questions concerning the world's food production, its energy requirements and the necessity of environmental policy, particularly to safeguard the earth's own resources. Other important problems might be brought forth. Not least with energy and environmental questions have the difficulties in relations between experts and laymen become highly current. Politicians must here often be regarded as laymen. For him who is neither expert nor politician, but has given some thought to the importance of a particular problem, there are two extremes to take: one relies entirely on the experts and politicians, or one troubles oneself to penetrate into the available facts, studies, and evaluations. Depending on abilities, some laymen may go even further and more or less become experts.

Should we leave the thinking to the experts? Would not that be a form of resignation, authority worship, or the easy way out — all rather antiquated in our day?

The need in each of us to be able to think critically has certainly never been so current as it is today. Many scandals such as the neurosedyne catastrophe in Sweden might have been avoided if awareness had been greater of the role of one's own thinking. The individuals entrusted with evaluating the drug chose to execute their responsibility by relying on studies done abroad and assurances given there, instead of considering the question as their own task.

Everywhere in society we find the need to be able to delve into accounts, reports, papers, studies, etc. The task often concerns not only learning the report's contents but perhaps even more importantly reading between the lines and trying to evaluate the author's motives, values, possible subjectivity, etc., to the extent that he hasn't accounted for them openly and honestly in the text. In popular presentations and in school books, facts are often mixed together with interpretations, the point of departure and the goals are only hinted at or not mentioned at all.

What is required here is the ability to analyze and listen. The Swedish school board's recommendation that education shall contribute to development of alert and logical thinking is more appropriate now than ever before. Development of thinking embraces, however, much more than the ability to criticize; for that mater, even much more than being able to think logically. The stronger and more important demand for creativity in the schools must naturally include even creativity in thinking. There is a need for creative fantasy everywhere, and when criticism is expressed, it ought always to suggest something constructive which can replace or improve the object being criticized.

Mathematics with its strongly contoured concepts, clearly delimited problem areas, and meager demands for materials. has the potential more than any other school subject for giving the pupils valuable development in creative and logical thinking.

The 2000 year old inheritance from th logical school in Greece has been invaluable, but during our century it has also been the cause of rigidity in forms of thinking. Liberation from the demonstrative method of proof, which goes back to Euclid and other Greeks, must continue within education.

In recent years there has developed a strong interest in the broad contributions to the field of mathematical heuristics made by the mathematician Georg Pólya. Two volumes were published in the United States by Pólya in 1954 on "Mathematics and Plausible Reasoning" (vol. I "Induction and Analogy in Mathematics", vol. II "Patterns of Plausible Inference"). Further work on problem-solving were published in 1961 and 1965.

Pólya has long been recognized as an active mathematician, but also as the author of pedagogically selected collections of problems, "Aufgaben und Lehrsätze aus der Analysis," which he wrote together

with G. Szegö. Pólya trained teachers in mathematics for more than 20 years. The books mentioned here have been translated to German as well as other languages.

A predecessor to these books was published in 1945 as a pocket book with the title *How to Solve It*. In a thoughtful review of the Swedish version in 1975, Dag Prawitz, professor of Philosophy at Oslo University, calls heuristics "a woefully ignored field."

> All that work which leads to the idea upon which a proof is based has no official status in the mathematical literature and is usually not mentioned at all. Modern school mathematics would look very different if consideration had been given to this field instead of so singularly concentrating on the logical part within the philosophy of mathematics.

Let us consider a few of Pólya's words:

> Mathematics is regarded as a demonstrative science. Yet this is only one of its aspects. Finished mathematics presented in a finished form appears as purely demonstrative, consisting of proofs only. Yet mathematics in the making resembles any other human knowledge in the making. You have to guess a mathematical theorem before you prove it; you have to guess the idea of the proof before you carry through the details. You have to combine observations and follow analogies; you have to try and try again. The result of the mathematician's creative work is demonstrative reasoning, a proof; but the proof is discovered by plausible reasoning, by guessing. If the learning of mathematics reflects to any degree the invention of mathematics, it must have a place for guessing, for plausible inference.
>
> The general or amateur student should also get a taste of demonstrative reasoning: he may have little opportunity to use it directly, but he should acquire a standard with which he can compare alleged evidence of all sorts aimed at him in modern life. But in all his endeavors he will need plausible reasoning. At any rate, an ambitious student of mathematics, whatever his further interests may be, should try to learn both kinds of reasoning, demonstrative and plausible.
>
> (From the preface of "Mathematics and Plausible Reasoning", vol.I, 1954)

Pólya does not mean to imply that there exists an infallible method for learning the art of guessing. But he addresses "teachers of mathematics at all levels" when he says, "Let us teach guessing."

Pólya's works contain a large number of examples from his experience as university teacher and mathematician, with which he illustrates how fantasy, trial, and investigation can proceed forward to reach the solution. The books are directed primarily to students of mathematics, and the examples are often taken from first year university courses but correspond to high school level in other cases. They have exceptionally much to give to the subject teacher of mathematics, especially to those who have a few years experience teaching at different levels.

The following questions should have priority when revising the mathematics teaching plan:

1. How to develop fantasy, the ability to get ideas, to guess?
2. How can the students learn to use experience they have gained?
3. How can they train self-discipline concerning logical thinking?

> More valuable than any particular mathematical fact or trick, theorem, or technique, is for the student to learn two things:
>
> First, to distinguish a valid demonstration from an invalid attempt, a proof from a guess.
>
> Second, to distinguish a more reasonable guess from a less reasonable guess.
>
> (From "Mathematics and Plausible Reasoning," vol. II, Chap. 16, Section 9.)

For years the high school curriculum has been designed basically with continued education at the university, teachers colleges, and other institutions in mind. Those recipients played a large role in the development of the 1970 Swedish School Curriculum Plan, even if in principle the school was to be organized such that the lower level would be independent of the higher levels.

The question now is if a change in mathematics curriculum in the direction of Pólya's guidelines would make high school students less capable of carrying on advanced studies. It is quite possible that what is lost in special knowledge and technique is more than compensated for by an ability in constructive thinking, thinking which would be more independent than what appears to be the case today.

Before we go on to examples and experience from our teaching, I would like to give here a few suggestions as to the direction which answers to the above three questions take:

1. The student must be given sufficient time to understand the problem itself, as presented. The first and immediate understanding directly following presentation of a problem is usually rather superficial. It ought to be deepened through questions, examples, and answers. The second understanding is much more likely to motivate the students. After this there must be time allowed for the students to dig into and explore the mathematical material offered – integers, proportions, triangles, or whatever it may be. They ought to seek personal experience, or best of all, their own discoveries through individual or group work. It is always exciting to go exploring.

2. Students should be given opportunity to vary the problems, simplify, choose special cases, generalize, or in other ways study related variants. By all means let them formulate and investigate problems of their own, choosing a problem of a similar nature. It is especially interesting trying to reach goals one has chosen for oneself.

(Once a girl in the 9th grade was indefatigable in carrying out divisions to investigate the length of the period in decimal fractions. The interesting thing is that she got going on this without my or the other pupils' being aware of it until she happily presented her results to us. The class was working on other things at the time.)

Results and partial results should be noted down in some kind of notebook so that comparisons can be easily made. It is well known that Michael Faraday carefully kept accurate and complete diary records of all his experiments. Even before he began his career as a researcher, he had taken copious notes and made detailed drawings at lectures he attended.

3. Development of self-control in logical thinking should be integrated with other subjects, particularly physics, chemistry, and English. It is a question here of observing one's own thinking and formulating it in words, activities which also are a part of the natural sciences and English.

A part of self-control includes checking, both while solving the problem and after reaching the solution, if all of the facts and conditions given in the text have been used. Problems with unnecessary, excess data occur. In applied mathematics it is often a part of problem formulation first to sift out those facts which are not relevant to the problem at hand. Much practice should be done describing what one has seen or thought. This gives good training in the objectivity and concentration which mathematics requires.

Terminology should be reduced to a minimum but consistently followed.

The pedagogical problem lies to a great extent in the question of how to awaken the pupils' interest. The methods which teachers in the lower classes use play a large role later on in the upper classes when the mathematics teacher takes over. If the lower class teacher feels enthusiasm when confronting problems and for different ways of approaching them, then this will rub off on the pupils. The more the teacher follows old routines, the less interest will be generated in the pupils. Cheap tricks to entertain the students are not necessary at all. "Bingo" games need not be used. When one of the textbooks for the "new math" came out here during the 1970's, it lacked a pedagogical method for the introduction of geometry. The children (in the 4th grade) were expected to entertain themselves by drawing people figures using circles and triangles, i.e. to do exercises which were entirely lacking in both artistic and mathematical content.

Many years' experience refutes the belief that motivation must be awakened with practical problems. Pupils can, of course, often advantageously be enthused through everyday occurrence, but they can and should be enthused even with purely mathematical questioning, as the next chapter will illustrate.

We shall return in Chapter 4 to the question of mathematics as a practice-arena for the thinking of young pupils after we have looked at teaching experiences and examples in the following chapter. The first section of Chapter 6, 6.1., and Chapter 8 discuss the introductory, orientation aspect of teaching mathematics.

3

THEMES FROM THE CLASSROOM

3.1 "How Many Are There?" — Numbers and Number Systems

3.1.1 How Many Are There?

Most children get well acquainted with our numbers in a natural fashion before they come to school. We need not and should not exercise them in counting. Children meet numbers in many games and otherwise in various daily situations, and eventually learn to experience them. They know that they have one nose, two eyes; they can see that three loaves are in the oven; they learn to identify the numbers 1 to 6 with the dot figures on dice, and so on. They may be able to count to 7, or perhaps to 17, when they start school. Some children find it exciting to count how many there are of various things — a joy which can grow during the first years of school. It can be a "sport" for a child to count the number of cars in a train passing by. They are not difficult to count when it is a passenger train; the cars are long enough to allow counting at a comfortable speed. But when we see an ore train pass by at a distance, it can be really hard. We must concentrate, strain ourselves to be as attentive as we possibly can in order not to lose count. There are so many cars and exactly alike. The difficulty seems to lie in keeping the cars separate from each other while counting. We need to "stop" them and retain individuality of each car as it passes by. Our glance must be strictly controlled by a determined will. When we count a pile of oranges, we often lay each one aside as we count it, but the ore cars cannot be put off to the side. Our whole awareness must be concentrated on our seeing.

This ore-train example shows that counting is always an act of will, and in actual fact when we count, we are counting our will impulses.

These impulses are so weak in most cases that the counting seems to go completely automatically.

In school we learn how to write the different numbers. We get more and more acquainted with the decimal system, learning its notation and arithmetic procedures. We become so used to our number system that we do not get any perspective on its elegant effectiveness until we are given the chance to build up another number system. Should we lose our memory and suddenly only be able to count to ten, then we would be troubled to construct a new system — very likely a 10-system — and we would appreciate it since we had constructed it through our own efforts.

During the 1st through 8th grades children have practiced the 10-system properly, so that arithmetic goes mechanically. One should not need to stop and think, for example, when given 25 · 35 to multiply. One works entirely within the system, although in fact quite without thinking, according to what one has learned earlier.

It is then of value to put oneself in another situation; for example, why not the Stone Age? We imagine that we can only count five stones or five arrows, etc. The concepts "six", "seven", etc. we do not have. How can we then describe the number of animals or arrows when that number is more than 5? Assume there are 38 arrows. How can we state that number when we only can count to 5?

The class will certainly have suggestions, for example, that we can draw a symbol for number 5, perhaps a hand (stylized) and state "units" up to four with slash marks:

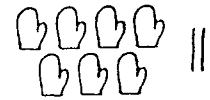
5 is written ; | || ||| |||| means 1, 2, 3, 4.

6 is given by , 7 by  etc.

How would, for example, 23 be written? Of course, as four hands and 3 lines:

And 38? As seven hands and 3 lines? No! No one knows how to count to 7, we are simply not able to grasp a group of 7 hands; we can only count to 5. What do we then do?

Our answer to this question will determine whether or not we eventually construct a more or less effective system. If we stop with only the two symbols – the hand and the slash — we could indicate 38 with this figure:



But what a lot of space this figure takes! And how much space would be needed for even larger numbers, assuming that we could learn to interpret them at all? What might help us toward a better system? We invent a new symbol, a symbol for 5 hands. In one ninth grade class they agreed upon using a shoe for symbol: in another they chose a five-pointed star. The picture for 38 would then be:

 or

How do we go on? Will any problems arise? It soon becomes apparent that a new symbol is needed for five shoes or five stars. Otherwise the same dilemma as earlier will occur. The above classes chose, respectively, a boot and a star with extended rays.

In the latter class there was another interesting suggestion for the symbols for 25 and 125 (compare Fig. 3.1.1):

 25  125

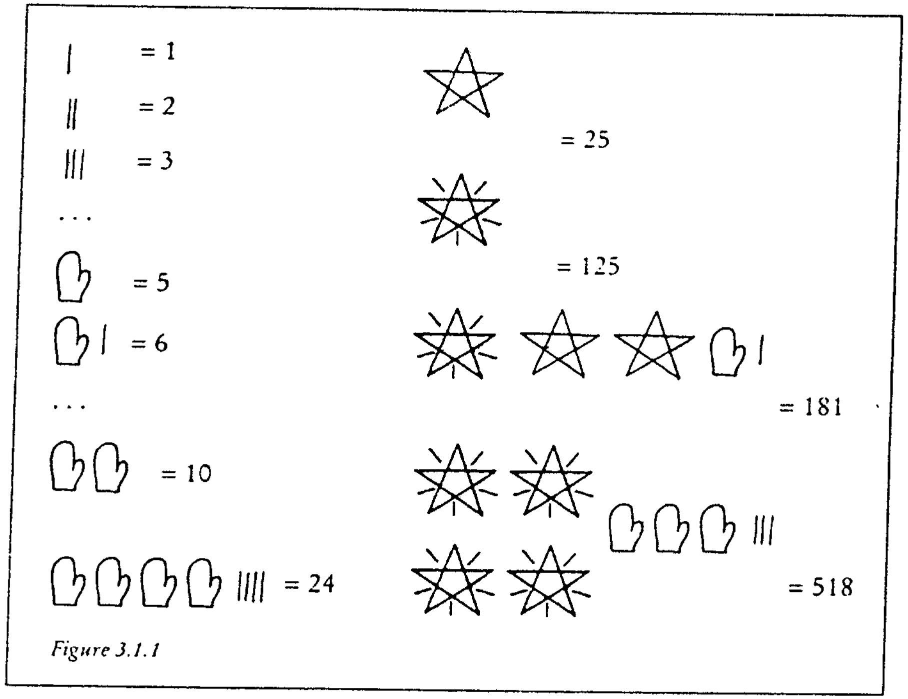
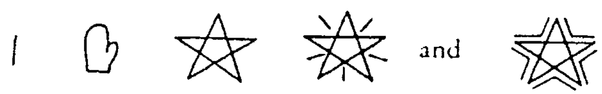


Figure 3.1.1

Extending is now easy. We have understood that the numbers 1, 5, 25, $5 \cdot 25 = 125$, $5 \cdot 125 = 625$, etc., build the base in our five-system. If we now restrict ourselves to the world of picture-symbols, then a separate symbol will be needed for each one of these numbers. Assume we have



for 1, 5, 25, 125, and up to 625. What is the largest number we will be able to write pictorially? Obviously

$$4 \cdot 625 + 4 \cdot 125 + 4 \cdot 25 + 4 \cdot 5 + 4 = 3124.$$

We get the number just below he next symbol number, 3125!

Now we should ask: can every number from 1 to 3124 be written unambiguously with our five symbols? Are there possibly one or more

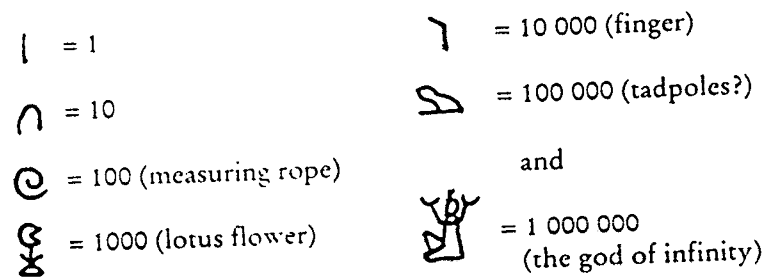
numbers which cannot be handled at all in our newly constructed five-system – or which can be written in more than one way?

It is not difficult to convince ourselves: our 5-system, extended indefinitely, gives unique representations of all the integers. (Exercise 3)

3.1.2 Number Systems in Early Cultures

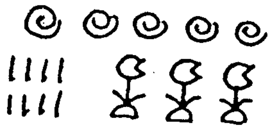
How did the ancient Egyptians construct their number systems?

Numbers and arithmetic operations were expressed, as was their normal writing, with picture-symbols. When numbers were inscribed on a relatively hard material, they used special hieroglyphs for the numbers. But when the arithmetic was done with Indian ink or a similar liquid on papyrus leaves, they wrote a relatively flowing style, so-called hieratic (holy) writing. Here we limit ourselves to the hieroglyphic writing. They used a 10-system where the base was composed of the following hieroglyphs:



The Egyptians used each symbol to represent as many units as the base number specified. To describe a number of cows one drew each individual symbol as many times as was required. For example, 3508 was written

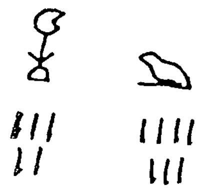
Since each symbol represented a certain fixed number, the symbol could be placed any which way. This too:



must mean 3508.

What corresponds to our zero? Is it needed? No, in this pictorial writing there is no need for a zero symbol.

The best documents showing us ancient Egyptian writing are the Rhind Papyrus and the Moscow Papyrus, kept at the British Museaum in London and in Moscow, respectively. Rhind is the name of the purchaser of the first-mentioned document. Both rolls have their roots in the epoch around 1800 B.C. in ancient Egypt. They contain a series of solved problems and could be called collections of recipes of the type "do this," "do that," "you will find it to be right." In a later period the Egyptians took the first step toward a more abstract way of writing numbers. One wrote out a sort of table where the columns correspond to different base numbers. For example, the number 705,000 would be written:



Strange as it may seem, the people of the Fertile Crescent around the Tigris-Euphrates valleys (the Akkadians, Babylonians, Sumerians, and Assyrians) used a completely different system for representing numbers. Wedge-shaped prism marks were stamped in soft clay (see Figure 3.1.2).

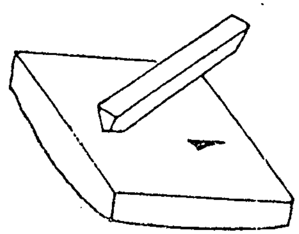


Figure 3.1.2
(After O. Neugebauer)

The writing is called cuneiform.

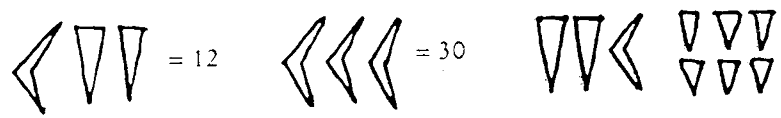
The base in the Babylonians' system was 60; the basic numbers were 1, 60, $60 \cdot 60 = 3600$, $60 \cdot 3600 = 216,000$, etc.

All of these numbers were indicated by the wedge-shaped stamp:

It was the context which showed the particular value the stamp had in particular place. In this system the number 59 would require 59 stamp marks, which would be difficult to read and understand. One would have to count the number of wedge marks and one could count wrongly, arriving at 58 instead of 59. They therefore had a helping symbol with the value of ten, a wing-shaped mark:

This figure corresponded to 10 wedges and thus could represent 10, $10 \cdot 60 = 600$, $10 \cdot 3600 = 36\,000$, $10 \cdot 216\,000 = 2\,160\,000$, etc.

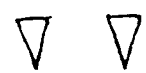
The following are some examples of number representations:



From these examples we see that the wedge's placement, its position, plays a determining role. For example, the wedge on the left could indicate the value 60 while the wedge on the right could mean 1. Here we have the beginning of a so-called positional system, a system in which the symbol's position determines which value that symbol adds to the number being represented. Our own system is a positional system.

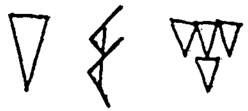
In the number 5358 both of the "5 s" represent 5, of course, but the left five contributes the value 5000 to the number while the right-hand five contributes 50. In this way we are able to use the same symbol to account for different values.

Was it always possible to determine from the context what the wedge marks mean? E.g. in



Could one decide whether the value was 2, 61, 120, or 3601?

From the fact that various clarifying symbols were introduced it seems clear that mistakes were made. One was to write a word beside the symbol. Another method was to use a sort of double-wing as a symbol for an "empty-space." For example:



$= 3600 + 0.60 + 4 = 3604.$

Symbols for the empty space are predecessors to the number zero. In ancient India the empty space was marked with a dot. Opinions are divided on when zero (naught) began to be used as a number, that is, on when zero was given the same status as the numbers 1, 2, 3, etc. According to van der Waerden it was the Greek astronomers who introduced the number zero, roughly around the time of Christ.

Where do our numbers come from?

We often say that we use arabic numerals, but the fact is that our number system came from ancient India. The individual forms of the numerals have in some cases undergone radical metamorphoses (Figure 3.1.3).

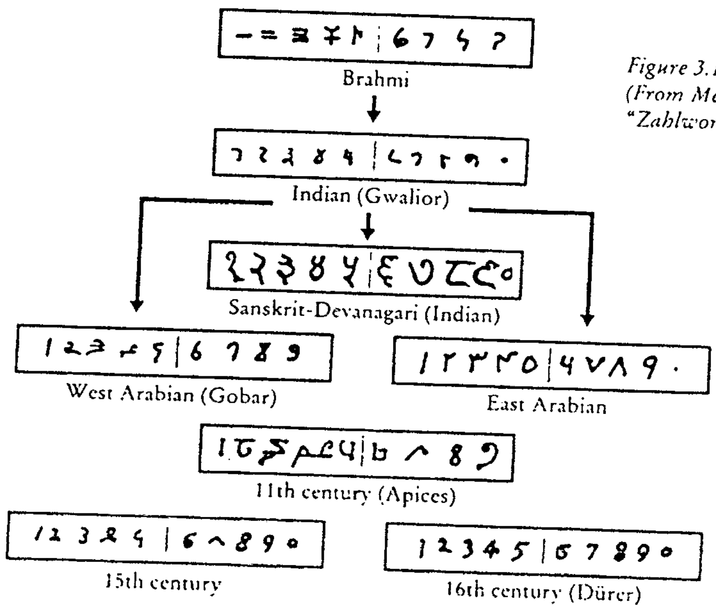


Figure 3.1.3
(From Menninger, "Zahlwort und Ziffer")

The Indian numerals were brought to the West by the Arabs, primarily during Islam's expansion westward in the first centuries after Mohammed's death. Albrecht Dürer (1471-1528) gave the numbers the forms which they by and large have retained to this day.

We have thus received our numerals from India and the number zero from Greece. The 10-system has its roots in old Egypt (and certainly further back in time) and the position-principle in its early form comes from the people of the Tigris-Euphrates valley.

Does the 10-system have connections to our 10 fingers (and toes)? Most certainly, and therefore ought its roots to be as old as man himself? Language research tells us:

- in German 10 is "zehn" and "toes" are "Zehen"
- in English "digit" means both finger and numeral (computers which give results as numbers are called digital computers – and in recent years we have digital clocks and watches)
- according to Tobias Dantzig, 10 is the base for the numbers in all the Indo-European languages and in some other languages as well

It may appear from the words "eleven" and "twelve" that we have had a twelve-system. Language research shows, however, that the German words for eleven and twelve, from which our English words come, "elf" and "zwölf" respectively, come from the compounds *ein-lif* and *zwo-lif*, where *lif* is an old Germanic word for ten.

In the following table we see some examples of number names from systems based on 20 and 5.

1. The Mayans used the system which you get if you choose 20 as your base, i.e. a 20-system. The day was divided into 20 hours. An army division was comprised of $20 \cdot 20 \cdot 20 = 8000$ soldiers and so on. Basic numbers:

1	hun	=	1
20	kal	=	20
20^2	bak	=	400
20^3	pic	=	8 000
20^4	calab	=	160 000
20^5	kinchel	=	3 200 000
20^6	alce	=	64 000 000

2. Typical primitive base-5-systems have been found with the Ap[i] people in the New Hebrides (the island group east of Australia) as we[ll] as with the Wanyassa tribe in Central Africa. (The information on th[e] Wanyassas stems from Stanley.) We can note that the systems are struc[-]tured in the same way, despite the great distance between these two cul[-]tures!

	Api culture	*Wanyassa*
1	tai	kimodzi
2	lua	vioviri
3	tolu	vitatu
4	vari	vinyé
5	luna	visiano
6	otai	visiano na kimodzi
7	olua	visiano na vioviri
8	otolu	visiano na vitatu
9	ovari	visiano na vinyé
10	lua luna	visiano na visiano

Luna means both five and hand in the Api language. The prefix "o" means "more", e.g. otai = "more one" or "one more", i.e. 1 + 5.

3.1.3 *The Five-System with Numbers*

We constructed earlier a five-system with the help of symbols for the basic numbers, i.e. with pictures for the numbers 1, 5, 25, 125, etc.

What do base-5-numbers look like if we refrain from our pictures and use our numerals instead? Can we make a table from right to left with columns in the order for 1, 5, 25, 125, ... just as in our 10-system with its columns for 1, 10, 100, 1000, etc.?

The numbers 1, 2, 3, and 4 can be written unchanged. But 5 must be written as the sum of 1 "five" and 0 "ones":

125	25	5	1
		1	0

5 is therefore written 10 in the five-system. To avoid confusion we attach little number tags to specify which number system we mean:

(Five in the ten-system is written "one-zero" in the five-system. We shouldn't read "one-zero" as "ten" because "ten" already has its meaning for us within the decimal system.)

In the same way we get:

$$6_{10} = 11_5$$
$$7_{10} = 12_5$$
$$\dots$$
$$10_{10} = 20_5$$
$$11_{10} = 21_5$$
etc.
$$25_{10} = 100_5$$
$$26_{10} = 101_5 \text{ etc.}$$

Those who wish to give a try on their own may do some of the exercises following. From the system's structure it is clear that no digit higher than 4 should appear – nor is there any need. The system has the five digits 0, 1, 2, 3, and 4. Adding the numbers 6 and 8 with 5-system addition gives:

11
<u>13</u>
24 (with the value fourteen)

Adding instead 78 to 113 we get with the aid of a table:

	125	25	5	1
78		3	0	3
113		4	2	3
	1	2	3	1

Here we begin with 3 + 3 getting the sum 11_5, in which the left-most one specifies a "five" and is carried over to the fives-column. This is why the fives-column adds up to 3 (1+0+2). Finally, the sum 7 in the 25-column gives a carry-over 1 in the 125s-column.

Ninth graders usually don't mind trying out the other three arithmetic operations in the fives-system. Those who are interested in seeing how these work out can take one one or another of the exercises at the end of the section.

3.1.4 The Binary System

By this time the pupils have asked which system computers use. There are computers built for the 10-system, but the majority of them as well as pocket calculators are based on the 2-system, or as it is called, the binary system ("bi" = two).[1]

Which basic numbers do we get if we choose 2 as our base?

We get 1, 2, 4 = 2 · 2, 8 = 2 · 4, 16 = 2 · 8, etc., in other words: a sequence starting with 1 in which each number doubles the previous number on and on.

The table here shows the numbers from 1 to 17 "translated" into binary language. Here too avenues open up for discovering on one's own how the operations addition, subtraction, multiplication and division work out.

	16	8	4	2	1
1 =					1
2 =				1	0
3 =				1	1
4 =			1	0	0
5 =			1	0	1
6 =			1	1	0
7 =			1	1	1
8 =		1	0	0	0
9 =		1	0	0	1
10 =		1	0	1	0
11 =		1	0	1	1
12 =		1	1	0	0
13 =		1	1	0	1
14 =		1	1	1	0
15 =		1	1	1	1
16 =	1	0	0	0	0
17 =	1	0	0	0	1

Why are computers based on the binary system? There might be a student in the class who wishes to prepare a sufficiently detailed yet easily understood answer to this question to present later to the class.

---

[1] Space does not allow here going into *how* the 2-system is used but examples in the classroom teaching may be to advantage.

Pupils should understand at the least that it is the basic 2-way polarity of electricity and magnetism and their phenomena which makes the 2-system so natural for technical calculation. The 2-system has, of course, only the digits 0 and 1. Corresponding to these digits could be: high and low voltage respectively, current flow versus no-current, or clockwise magnetization versus counter-clockwise magnetization.

We come naturally into a discussion of the advantages and disadvantages of binary numbers compared to decimal (the 10-system). Look how easy the multiplication table is in the two-system:

$$0 \cdot 0 = 0$$
$$0 \cdot 1 = 0$$
$$1 \cdot 0 = 0$$
$$1 \cdot 1 = 1$$

But, on the other hand, — oh, such long numbers! The number one hundred must be written 1100100 in the 2-system. Doing arithmetic in the 2-system with pen and paper would be, in fact, a trouble, but with the incredibly fast computers it doesn't particularly matter. An addition of 2 digits can be done in approximately 60 billionths of a second.

Why work with this material in grade 9? We take up this question in Chapter 7 (Section 7.2). See also section 3.2.7.

To conclude this orientation on number systems we will take a look at two primitive binary systems.

Similar primitive base-2-systems have been found as far apart as in a tribe at Torres Straits, Australia, and in the Bakairi tribe in central Brazil:

	At Torees Straits	*In the Bakairi tribe*
1	urapun	tokale
2	okosa	ahage
3	okosa-urapun	ahage-tokale
4	okosa-okosa	agahe-ahage
5	okosa-okosa-urapun	ahage-ahage-tokale
6	okosa-okosa-okosa	ahage-ahage-ahage

For numbers larger than 6 the Bakairis have only the word "more," which means "many."

Again, we have an example of two systems from completely diffe ent parts of the world which despite the distance are the same wi respect to the structure of the words for numbers. In what way are the primitive? The answer comes directly from the words themselves; th system is structured additively: one, one-plus-one, two-one, two-two etc. No wonder that they didn't get very far along this route. The num ber sixteen would sound (in English):

two - two - two - two - two - two - two - two !

The binary system which our culture uses is based upon the *multi- plicative* principle, in a similar fashion to our decimal system and also to the 5-system we constructed earlier. After 1 and 2 we do not go to 1 plus 2 but to 2 "times" 2 = 4 and after that to 2 · 4 = 8 etc. Multiplication by 2 repeats, and the basic numbers are what we call the powers of 2.

2 is the first power of 2 and is written 2^1
4 is the second power of 2 and is written 2^2 (meaning 2 · 2)
8 is the third power of of 2 and is written 2^3 (2 · 2 · 2) etc.

It is natural to complete this system with the definition $1 = 2^0$.

Among the Thimshian people in British Columbia have been found 7 different sets of number words. One set for flat objects and animals, one for round objects and time, one for counting people, a set for long objects and trees, a set for canoes, a set for measurements, and a set for counting when no particular objects are specified. The first six sets point to times long ago when this people's counting was very much sensory-bound. As T. Dantzig points out in his book, *Numbers — the Language of Science*: "The concrete preceded the abstract."

We have not touched here upon fractions in the different systems. At school we have to some extent gone into such questions as what kind of binary language would correspond to tenths, hundredths, etc., what symbols could we use for binary fractions, etc.? Here I will content myself with referring the interested reader to exercise 7. Quick pupils can be given the task of building up systems with bases larger than 10, for example a 12-system or a 16-system. The 12-system requires two new numerals for ten and eleven, while the 16-system requires characters for the six numbers 10-15. One can use letters, for example A = 10, B = 11, etc.

3.1.5 Figurated Numbers

During the later phases of the ancient Greek civilization, especially among the Pythagoreans (Pythagoras was active in Greece during a long period around 500 B.C.), the whole numbers were presented as figures, so-called figurated numbers. Triangle numbers, square numbers, rectangles, etc. were formed (Figure 3.1.4).

The square numbers are determined by the number of dots in the squares and are thus 1, 4, 9, 16, 25, etc., the result of squaring 1 · 1, 2 · 2, 3 · 3, 4 · 4, 5 · 5, etc. The rectangle numbers are 2, 6, 12,20, 32, ... (consider: the area of a rectangle is "length times width"). The eleventh rectangle number is apparently 11 · 12 = 132, the 26th would be 26 · 27 = 702, and the n-th would be n · (n + 1).

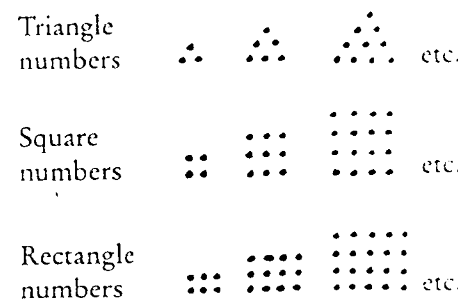


Figure 3.1.4

Now what values do we get for the triangle numbers?

In the beginning we have 1, 3, 6, 10, 15, ...

How does the series continue? Can we find a better method than drawing triangle after triangle and counting the dots?

In a 7th or 8th year class — or wherever else the question is raised — there will be no shortage of proposals. Some pupils look at the structure of the triangles and find that each triangle is made of rows: the top row has 1 dot, below it 2 dots, and so on with each lower row having one more dot until the last row which has as many dots as the number of the triangle (its order in the series of triangles). Thus we should write, for example,

1 + 2 + 3 + ... + 11 for the eleventh triangle number,

call it t_{11}, and calculate the sum.

Other pupils have examined only the numbers themselves, from the first few triangles:

$$t_1 = 1$$
$$t_2 = 3$$
$$t_3 = 6$$
$$t_4 = 10$$
$$t_5 = 15$$

and found that the increases from triangle to triangle are 1, 2, 3, 4, 5. They say "the next increase will be 6, the following 7, etc." Can we be sure of this? "Yes," say those who have studied the triangles. "It's right, because for each step we add a new bottom row which has one dot more than the bottom row of the previous triangle." So in principle we can work out any triangle number we want to, for example

$$t_{100} = 1 + 2 + 3 + \ldots + 100.$$

If it hasn't been discussed earlier one may here take up the question of how sums of this type, of so-called arithmetic series, are calculated. And with that, tell the classical story of how Gauss as a 10-year-old schoolboy solved a similar, although probably more difficult, problem. In such a way we can obtain a formula which directly calculates the triangle number without adding up a sum.

Is there another way, a way to directly calculate the triangle numbers? Can we find a geometrical solution by putting together two dot-triangles? Perhaps a little hint is needed. Can we put two identical triangles together? Of course, making a parallelogram (Figure 3.1.5a).

Figure 3.1.5a

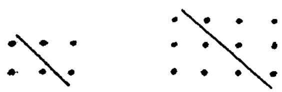


Figure 3.1.5b

And the parallelograms can be judiciously shifted into rectangles – without changing the number of dots (Fig. 3.1.5b).

Now we can see the triangle numbers as half of the corresponding rectangle numbers. We thus get, for example,

$$t_5 = \frac{5 \cdot 6}{2} = 15 \qquad \text{and} \qquad t_{100} = \frac{100 \cdot 101}{2} = 5050.$$

And the general formula for the triangular number is

$$t_n = \frac{n\,(n + 1)}{2}.$$

Let us look back upon the different levels in our calculations of the triangular numbers:

1st level: we draw the triangle and count its dots

2nd level: we note the triangle structure and form a sum to be calculated. We no longer need draw the triangle. The crucial point is that we identify the base row's length with the number of the triangle (its order in the series). From this we draw the conclusion that the sum includes all the integers from 1 up to and including the number of the triangle. We have found a general method.

3rd level: We wish to improve the method and find a geometrical solution which directly gives us the triangle number as half of the rectangular number of the same order. We can state a formula:

$$t_n = \frac{n\,(n + 1)}{2}.$$

All pupils can work at the "physical" level, and many can contribute with ideas and comments which lead to the other, truly mathematical, levels of solution.

This problem, in all its simplicity, gives much insight into how we can let geometrical clarity and experience with numbers lead us to a crucial idea.

3.1.6 Two Problems in Subdividing a Circle

In connection with the figurated numbers, the following two closely related problems are of interest:

1. A circle is subdivided into as many subdivisions as possible by 1 chord, 2 chords, 3 chords, etc. How many subdivisions are possible when the number of chords is

a) 4? b) 11? (Figure 3.1.6)

Start by making a table up to $n = 6$.

2. On the perimeter of a circle we place a point. With this point only we have just one subdivision, the whole circle. We place another point somewhere on the perimeter and draw the line connecting it to the

first point. Now the circle is divided by this chord into 2 areas. Place out a third point, a fourth, etc. in such a way that the number of subdivisions is maximal when all the points are connected by lines (all three points with each other, etc.). Make a table as in Figure 3.1.7 with the maximum number of subdivisions.

The first question is: what do we do, how can we satisfy the condition for maximum number of subdivisions?

A further task: find a formula, in Problem 1, for calculating the number of subdivisions (Exercise 8).

After we have acquainted ourselves with Problem 1 by drawing figures, a little thought brings us to the realization that we must place each new chord so that it intersects all of the old chords but at different points on the different chords.

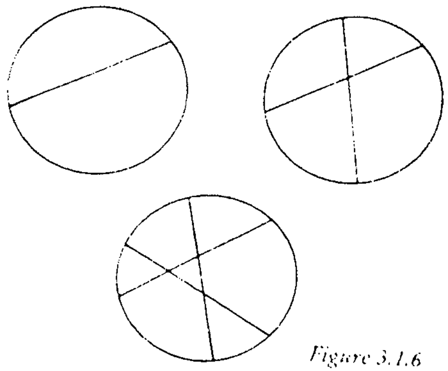


Figure 3.1.6

n Number of chords	A (n) Max. number of areas
1	2
2	4
3	7
4	?
5	?
6	?

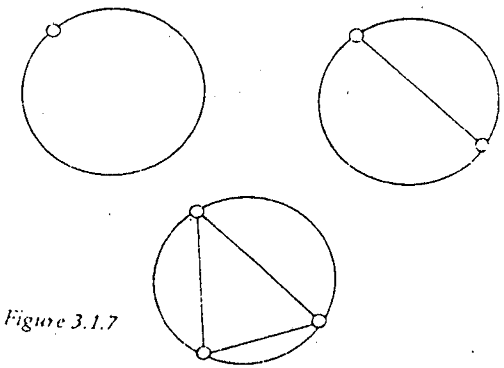


Figure 3.1.7

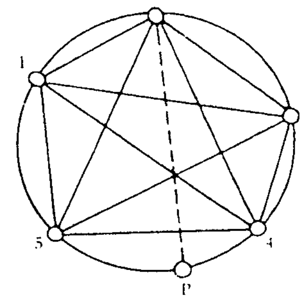
P Number of chords	A (p) Max. number of areas
1	1
2	2
3	4
4	?
5	?
6	?

We must not draw a new chord through an intersection of two old chords.

In Problem 2 we are not concerned with the maximum requirement until we come to the 6th point.

The sixth point must avoid positions (there are 5 possible) such that a connecting line to one of the old points goes through an intersection of two earlier chords (Figure 3.1.8).

Figure 3.1.8



Students generally come quickly to agreement on the results in Problem 1:

Number of chords	Max. number of areas
1	2
2	4
3	7
4	11
5	16
6	22

During work on Problem 2 opinions have to date always been divided. In one class the quickest pupil jumped up and shouted hotly:

"It has to be 32! It's obvious; it must be 32!" On other occasions pupils have worked quietly for themselves and found 31 subdivisions for 6 points, but believed it to be wrong ("It ought to be 32."). They made new figures. "Perhaps I didn't check off *all* the areas" is a common reflection. But it turns out, after carefully checking off the subdivisions, that the table looks like this:

Number of points	Max. number of areas
1	1
2	2
3	4
4	8
5	16
6	31

Many are surprised that the number of areas does not again double when we come to 6 points — the sequence previously was, to be sure 1, 2, 4, 8 and 16.

Before we had time to discuss the question further, some students have set out 7 points and counted 57 subdivisions. With 8 points the number rises to 99. The sequence thus diverges more and more from the sequence with successive doubling: 1, 2, 4, 8, 16, 32, 64, 128, ...

We will give a name to our new sequence – 1, 2, 4, 8, 16, 31, 57, 99, ... and return to it in a later section (3.2.6). We call it Moser's sequence, because the problem was formulated by Leo Moser (*Scientific American*, August, 1969).

Moser's example gives us useful experience: many times we are deceived by a preconceived idea about the results. The reason might be, as in our example, that we generalize too early, something for all analysts and decision-makers to give thought to.

3.1.7 Prime Number Generators

Often cited examples with surprising results are the following:

1. What values does the expression $x^2 + x + 41$ take on when we successively set x to 0, 1, 2, 3, ...?

Can we see anything in common among the values we obtain?

Putting in $x = 0$ gives $0^2 + 0 + 41 = 41$
$x = 1$: $1^2 + 1 + 41 = 43$
$x = 2$: $2^2 + 2 + 41 = 47$
$x = 3$: $3^2 + 3 + 41 = 53$
$x = 4$: $4^2 + 4 + 41 = 61$

Do 41, 43, 47, 53, and 61 have anything in common?

Yes, they are so-called *prime numbers* (a number n which is not divisible by an integer between 1 and n).

Is $x^2 + x + 41$ an expression which only generates prime numbers?

We test this idea by putting in $x = 5$: $5^2 + 5 + 41 = 71$.

A prime number!

Further testing can be easily divided up in the classroom, so that each pupil has his or her number to put into the expression. This will show that new prime numbers continually turn up. The next five are

for $x = 6$: 83
$x = 7$: 97
$x = 8$: 113
$x = 9$: 131
$x = 10$: 151

We see easily that some prime numbers are skipped over. Between 53 and 61 lies 59, between 83 and 97 lies 89, etc. But – does the expression always give prime numbers, even if there are many missing? Or are there values of x, which do not give prime numbers when substituted in the expression?

There are such x-numbers. Let's put in 41. What happens?

We get $41^2 + 41 + 41$ — a sum which is very obviously divisible by 41: the value can be re-written $41 \cdot 41 + 2 \cdot 41$ or more simply $43 \cdot 41$. We see that the number is composed of the two prime number factors 41 and 43.

Leonard Euler (1701-1783), from whom the example comes, found that all the 40 values of x from $x = 0$ to $x = 39$ give prime numbers. (It is not difficult to show that $x = 40$ gives a value which can be written $41 \cdot 41$, so that even $x = 40$ gives a composed value.)

Let us now form a spiral of squares starting with the number 41 (Figure 3.1.9).

Where do the prime numbers which we have obtained from the x-substitutions appear? (See Exercise 10.)

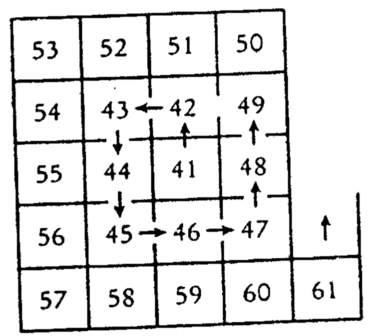


Figure 3.1.9

Euler also gave the expression $x^2 + x + 17$ as another surprising "prime number generator." Since the constant 41 is exchanged here for the lower number 17, the work with the quadratic spiral will be less demanding.

The expression $x^2 + x + 17$ is a so-called polynomial with one variable (x) and of the second degree (the highest exponent is 2, in the x^2-term). It has been proved that no polynomi can generate *only* prime numbers, not even if the polynomial contain more variables.

But there are polynomials of 12 variables which manage to generat all the prime numbers, along with some non-primes. A relatively simpl formula, (beyond, however, the scope of the school curriculum) was published in 1975 which gives the prime numbers p ($p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc.) successively when putting in the numbers n = 1, 2, 3, ... Interesting results have been published in the *American Mathematic Monthly*, vol. 83, 1976, and the *Canadian Mathematics Bulletin*, vol. 18 (3), 1975.

3.1.8 *Exercises*

1. Write the numbers 734 and 12059 with

   a) old Egyptian hieroglyphs
   b) cuneiform (Babylonian wedge-marks)

2. Write the numbers a) 39 b) 150 c) 795 in the fives-system. (Use the numerals 0, 1, 2, 3, 4.)

3. Can every natural number be represented in a fives-system? Uniquely?

4. Convert the following numbers to ordinary decimal form (to 10-system numbers):

   a) 34_5 b) 230_5 c) 304_5 d) 10110_2 e) 11011_2

5. Write out the multiplication table in the 5-system for numbers up to 4 · 4. In other words, finish the table below where a number in the left-most column is multiplied by a number in the top row and the result put in the appropriate place in the table. Some products are already filled in to get you started.

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	11	13
3					
4					

6. Carry out the following arithmetic in the given systems without using the 10-system:

   a) $31_5 + 33_5$ b) $10_2 + 101_2$ c) $10101_2 - 110_2$ d) $101_2 \cdot 11_2$

   Check, if you wish, by converting all numbers (both in the problem and in the answer) to decimal form.

7. Analogous to the decimal comma (decimal point) in the 10-system, we introduce a binary comma (;) in the 2-system and utilize the basic fractions

$$0;1 = \frac{1}{2} \qquad 0;01 = \frac{1}{4} \qquad 0;001 = \frac{1}{8} \quad \text{etc.}$$

   Do the following sum in the 2-system and convert the result to a decimal number: 1;01 + 0;011 + 0;101

8. Derive a formula for the maximum number of subdivisions in the first circle subdivision problem, where n = the number of chords (3.1.6).

9. Determine the values corresponding to the so-called pentagonal numbers as illustrated below in Fig. 3.1.10. (Try to work out a formula or at least an expression which gives the right values.)

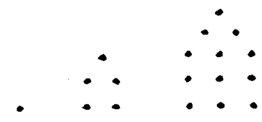


Figure 3.1.10

10. Referring to the Prime Number Generators earlier, can any of the numbers generated by the expression

$$x^2 + x + 41$$

land anywhere but in the corner boxes of the quadratic spiral?

3.2 Pascal's Triangle

3.2.1 Street Network

We continue on our voyage of discovery through the world of numbers and once again examine counting with a concrete example.

Figure 3.2.1 illustrates a street system in a modern city where avenues and boulevards form a network of squares. The avenues in the figure run downward to the right; the boulevards downward to the left. We can label the streets and the intersections. Corner H in the figure is the intersection of Avenue 2 and Boulevard 3 and may thus be called A2, B3. From the top corner AO, BO one may proceed to H by different paths. Some paths give a route of minimum length, namely 5 blocks long. Ignoring all longer paths we ask: how many paths of shortest possible length go from AO, BO to A2, B3?

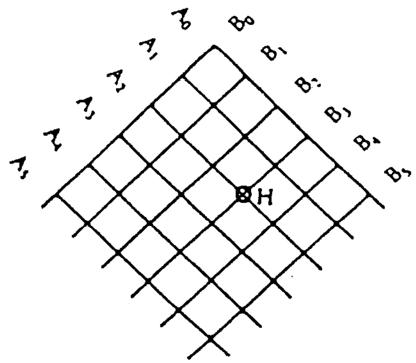


Figure 3.2.1

Many students choose to be concrete and draw the paths which they find. All pupils have the ability to approach the problem in this way. But soon the question arises: have we found *all* of the paths or is there yet another route left?

The pupils compare their figures, of course. Some have 8 routes, others 9, others 10 and someone has 11 routes. How can we *know* that we have found the last path? We need to *systematize!*

We need to order all the paths according to a system – when we can determine which route will be the last and in this way know the number of paths. How can we come upon an organizing principle such that we can number the paths and let the next path develop logically from the previous one?

Figure 3.2.2 shows a sequence of routes which some of the pupils have found. According to what principle are the routes ordered in the figure? The answer is given in Exercise 1.

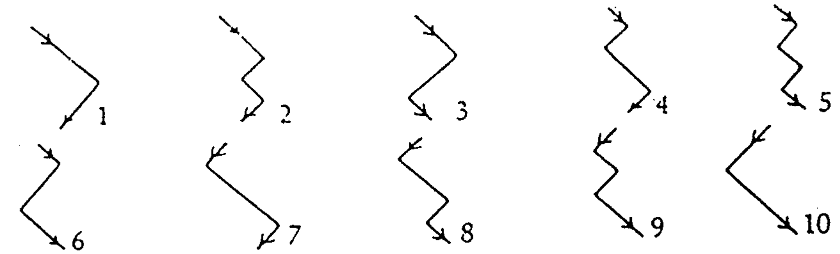


Figure 3.2.2 A systematic sequence of paths in Section 3.2.1 See also Exercise 1.

A problem such as this speaks naturally to the intellect but appeals also to the imagination. We can draw our street system and count successively how many shortest paths go from the top corners to different intersections. We start with the nearest intersections and write in the results on our drawing. To begin with we get, schematically:

```
  1   1
1   2   1
```

How does the number diagram continue to unfold?

Let us continue on until the shortest paths are of length 6 blocks.

Let us observe the numbers we get. Perhaps we may soon make a discovery and thereby come to a theory of how the numbers increase.

Perhaps we may even see directly during our work exactly how the numbers grow. That is, of course, our real goal.

We can in any case all help out in developing the number scheme. The correct scheme looks like this

```
              1   1
            1   2   1
          1   3   3   1
        1   4   6   4   1
      1   5  10  10   5   1
    1   6  15  20  15   6   1        etc.
```

Do these numbers show us anything interesting? Is there any regularity to be discovered? Doesn't it appear as if each number of paths is the sum of the nearest numbers in the row above? (The numbers on the edges are, of course, always 1.) We can check the sums in the entire diagram:

$3 = 1 + 2$, $4 = 1 + 3$ or $3 + 1$, $10 = 6 + 4$ or $4 + 6$ etc.

Does this rule apply generally? Could we, for example, prove that the number of paths to A2, B3 has to be 6 + 4, i.e. has to be the sum of the number of paths to A2, B2 and to A1, B3?

And could we thereafter realize that for any arbitrary choice of intersection:

$$z = x + y \text{ ? (see Figure 3.2.3)}$$

The chain of thought in a group – or individually – might go roughly like this: we keep in mind that only the shortest paths are to be considered. How must these look in the network? Well, all the shortest paths to the Z-intersection must pass through either intersection X or intersection Y. How many paths pass through these intersections? Obviously numbers x through X and y through Y respectively. The paths to corner X are simply extended by one block, and this does not change their number. "We just go one step further with each path," might be the pupils' expression. In the same way the number of paths to corner Z through corner Y is y. Thus

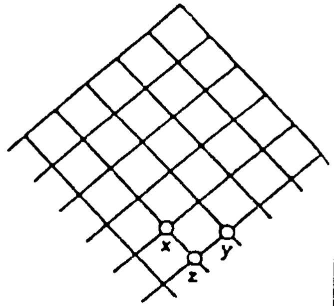


Figure 3.2.3

$$z = x + y$$

and we have proven an additive rule which allows us to further expand the number diagram.

We can leave behind the rather primitive and time-consuming method of looking for each new intersection, for all possible shortest paths and then counting them on our fingers, so to speak. Let us expand the table another 6 or 7 rows down.

```
                              1   1
                            1   2   1
                          1   3   3   1
                        1   4   6   4   1
                      1   5  10  10   5   1
                    1   6  15  20  15   6   1
                  1   7  21  35  35  21   7   1
                1   8  28  56  70  56  28   8   1
              1   9  36  84 126 126  84  36   9   1
            1  10  45 120 210 252 210 120  45  10   1
          1  11  55 165 330 462 462 330 165  55  11   1
        1  12  66 220 495 792 924 792 495 220  66  12   1
```


"Gosh, this is getting to be work. Isn't there any easier way to do this, Mr. Ulin?" "Isn't there some formula to directly calculate the numbers so that we don't have to add up the numbers in the table, row by row?"

We shall take up this question later but leave it to rest for a while. Now we turn our attention to...

3.2.2 A Completely Different Problem

Of five people, whom we shall call A, B, C, D, and E, we wish to choose two to do a task. How many pairs can be formed from the five persons? How many combinations of two objects are there from five different given objects?

One can approach the problem in many ways. Some pupils write down all combinations systematically in the alphabetical order:

AB	BC	CD	DE
AC	BD	CE	
AD	BE		
AE			

and arrive at the sum 4 + 3 + 2 + 1, just as in the study of triangle numbers (Section 3.1).

Someone says: if we write down all 4 combinations for each person, then the table will look like this:

AB	BA	CA	DA	EA
AC	BC	CB	DB	EB
AD	BD	CD	DC	EC
AE	BE	CE	DE	ED

Here every combination appears accounted for twice, for example, BA in addition to AB. The number of combinations is therefore

$$\frac{5.4}{2}$$

This method gives directly the sum to which the previous method led.

We can also let the five people be represented by 5 points, regularly arranged as the corners of a pentagon, and seek the number of possible connecting lines between them. Each connecting line then corresponds geometrically to one combination of two people.

If the number of people is n, then the number of choices of two persons — according to the reasoning which just gave us $\frac{5 \cdot 4}{2}$ as the number in the case of 5 people — will be

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

We introduce here the nomenclature $\binom{n}{k}$ for the number of combinations of k people from among n people, or more generally, k objects from n different objects.

$\binom{n}{2}$ gives us the Greek triangle numbers:

$$\binom{2}{2} = 1 \quad \binom{3}{2} = 3 \quad \binom{4}{2} = 6 \quad \binom{5}{2} = 10 \quad \binom{6}{2} = 15 \text{ etc.}$$

A look at the number diagram for the street network earlier shows that the numbers there form two symmetrical lines in the triangular table. For example, the number of paths to those intersections lying along Boulevard 2 give us the triangle numbers 1, 3, 6, 10, 15, ... This leads us to ask:

Can we find an equivalence between the street network problem and the combinatorial problem? Are these problems only different in *appearance;* are we here basically concerned with one and the same mathematical problem?

To form a combination of two persons from five means that I must choose two people. Do I encounter any choices when I go from AO, BO to the A3, B2, which, as we remember, has 10 shortest paths to it?

Of course! Right at the start I choose to go left or right (defined, let us say, from the reader's point of view). I will pass another 4 intersections before I arrive at A3, B2. In order to get there I must choose left at two of the intersections. *Where* I choose to turn left does not matter, just that I turn left exactly two times. The situation is the same when choosing two people. I can go to each one of them and say "yes" (selecting them) or "no." Exactly twice must I say "yes" if I am to have two people in my combination.

In this way it is clear that the number of paths from the starting point to the intersection A7, B5, for example, is the same as the number

of combinations of 7 (or 5) people from among 12 people. The number of paths to the intersection will thereby be:

$$\binom{12}{5} = \binom{12}{7}.$$

(This equality is understood at once when we consider that "yes" to 5 people means "no" to the other 7 people.)

3.2.3 Pascal's Triangle

We can now take up the question whether there exists a direct formula for the numbers $\binom{n}{k}$ which constitute the number scheme, or Pascal's triangle, as the scheme came to be called after the French mathematician and philosopher Blaise Pascal (1623-1662). Pascal wrote a very stimulating dissertation on the scheme in 1653: "Traité du triangle arithmetique."

This arithmetic triangle was known before Pascal's time, however. Documents have shown that the scheme existed in China early in the 1300's. The triangle begins with a one at the top and thereafter is identical with the scheme for the number paths in the street network (Section 3.2.1).

```
                1
              1   1
            1   2   1
          1   3   3   1
        1   4   6   4   1
      1   5  10  10   5   1
  . . . . . . . . . . . . . . . .
```

If we wish, we may re-write the triangle as below, where we let $\binom{n}{0}$ have the value 1, just as we do for all the natural numbers n:

$$\begin{array}{ccccccc} & & & \binom{0}{0} & & & \\ & & \binom{1}{0} & & \binom{1}{1} & & \\ & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & \\ \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} \\ \end{array}$$

. . . .

Can the $\binom{n}{k}$ -values be calculated directly? When pupils construc Pascal's triangle, they find the additions rather laborious as they com down to rows with 2-digit numbers. Tables may exist for $\binom{15}{k}$ or perhap even a lower row, but how would you determine $\binom{52}{13}$, for example?

Doing it by addition is far too time-consuming and unsure, even with the help of a pocket calculator.

Direct calculation of $\binom{n}{k}$

Let's take $\binom{9}{4}$ as a concrete example the number of combinations of 4 persons from 9, whom we could call A, B, C, ... H and I.

If we start by trying to follow the method we used for $\binom{5}{3}$, we would begin by asking: How many groups of letters are there with A as the first letter? When A is the first letter, how many choices have we for the second letter? There are 8 choices. And after that, when the second letter is chosen, in how many ways can we select letter number 3? There are 7 choices. And so on. The number of choices possible for letter number 1, 2, 3, etc. is 9, 8, 7, ... respectively down to 2 and 1. "One choice" for the last letter must here be interpreted as meaning that we have to choose the letter remaining. (Some students argue here that we have no choice at all....)

We may illustrate the number of choices of the successive letters with a kind of tree or street network which subdivides and spreads out more and more in all directions (Figure 3.2.4).

The figure illustrates the choice of the second and third letters and we can easily imagine for ourselves this network expanded to include the choice of the 4th letter. The number of groups with A as the first letter must be $8 \cdot 7 \cdot 6$.

Since the first letter may be chosen in 9 ways, the total number of 4-letter groupings must be $9 \cdot 8 \cdot 7 \cdot 6$ (= 3024)

But wait a minute! Do not the four letters A, B, C, and D occur as a group several times among these 3024 groupings? For example,

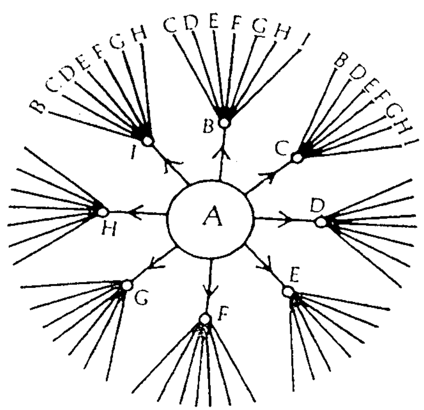


Figure 3.2.4

ABCD, BCAD, CDBA, DBCA, etc., which differ only in their ordering, must be considered as one and the same group of people, as only *one* combination of four elements. How many times do persons A, B, C and D occur together among the 3024 groupings which were formed when we considered the placement of the letters, when we ranked, so to speak, the people into first, second, third, and fourth places?

A, B, C, and D occur as many times together as there are possible placements of these four letters.

How many different placings are there?

For the first position in a 4-letter group we have 4 choices, for the second position 3 choices, for the third position 2 choices and for the fourth position "1 choice." This leads us to the number of different groupings, or permutations as they are usually called: the number is

$$4 \cdot 3 \cdot 2 \cdot 1 = 24.$$

This is how the list of 24 groupings looks, in alphabetical order:

ABCD	BACD	CABD	DABC
ABDC	BADC	CADB	DACB
ACBD	BADC	CADB	DACB
ACDB	BCDA	CBDA	DBCA
ADBC	BDAC	CDAB	DCAB
ADCB	BDCA	CDBA	DCBA

To summarize: of the 9 letters A through I we form $9 \cdot 8 \cdot 7 \cdot 6$ four-letter groups, where each letter elected occurs only once (AABC, for example, is not allowed). But each group of four letters occurs 24 times. The number of combinations must therefore be

$$\frac{9 \cdot 8 \cdot 7 \cdot 6}{24}$$

or more clearly written,

$$\frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2 \cdot 1}$$

which, when evaluated, gives the number 126.

The logic we have followed has shown us that

$$\binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2 \cdot 1}$$

and could be applied to any other particular numerical example without change. Thanks to the explicit appearance of each factor in expression (1), we can see through the particular case and grasp the general. How would we write, for example, the corresponding expression for $\binom{17}{6}$?

The answer is clear to us in a moment as soon as we become aware of how many factors there should be in the number (above the line); in the new example there are 6 objects to be chosen. We thus get

$$\binom{17}{6} = \frac{17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$$

The students take pleasure in reducing factors top and bottom and finally arrive at the value 12376. If the example were $\binom{17}{11}$, we would get in the same way

$$\binom{17}{11} = \frac{17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7}{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$$

which can be simplified at once to expression (2) for $\binom{17}{6}$.

Considering the fact that selecting 11 objects from among 17 different objects is equivalent to choosing 6 to throw out, we might directly have made use of the equality

$$\binom{17}{11} = \binom{17}{6}$$

The general formula becomes

$$\binom{n}{k} = \frac{n(n-1)(n-2)\ldots\{n-(k-1)\}}{k(k-1)(k-2)\ldots 2 \cdot 1}$$

or more simply

$$\binom{n}{k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k(k-1)(k-2)\ldots 2 \cdot 1}$$

We note that the number of permutations of n different objects is

$$p_n = n\,(n-1)\,(n-2)\ldots 2 \cdot 1$$

The expression for is usually written n! and called "n factorial." For example:

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$

(the 1 may of course be left out, if one wishes.)

The formulas for p_n and $\binom{n}{k}$ can then be written

$$p_n = n!$$

See also Exercises 3-5 below.

3.2.4 *Binomial Coefficients*

In numerical tables and in the mathematical literature in general, the numbers $\binom{n}{k}$ are called *binomial coefficients*. Where does this name come from?

A binomial is an expression which is formed as the sum of two terms, e.g. a + b, a + x, 1 + x, etc. Oftentimes we need to raise such binomials to powers, for example:

second powers $(a + b)^2$, $(1 + x)^2$ etc.
third powers $(a + b)^3$, $(1 + x)^3$ etc.

For example, if the side of a cube having length 1 is increased by x units, the new volume will be the third power of the new side, i.e.

$$\text{volume} = (1 + x)^3$$

Which terms appear if we expand these power expressions?

We perhaps remember the squaring rule which says that

$$(a + b)^2 = a^2 + 2ab + b^2.$$

The coefficients, as they are called, for a^2, ab and b^2 are here 1, 2 and 1 respectively.

(The expression may, of course, be written $1a^2 + 2ab + 1b^2$.)

Those who don't remember the squaring rule could always write (a + b) (a + b) for $(a + b)^2$ and carry out the multiplications step by step:

$$a(a + b) + b(a + b) = a^2 + ab + ab + b^2 = a^2 + 2ab + b^2.$$

Analogously we may write

$$(a + b)^3 = (a + b)(a + b)(a + b).$$

Here we need to form and add up all products which have exactly one factor from each of the parentheses. What kinds of products can appear?

Three a-factors gives $a \cdot a \cdot a = a^3$
Two a-factors and one b-factor gives $a \cdot a \cdot b = a^2b$
One a-factor and two b-factors gives $a \cdot b \cdot b = ab^2$
and finally three b-factors gives $b \cdot b \cdot b = b^3$

But how many products do we get of each kind?

Obviously one a^3 and one b^3.

In order to get an a^2b product, we need to choose b once out of three choices (for each parentheses we have the choice of a or b). The number of such choices (combinations) is

$$\binom{3}{1} = 3$$

The term ab^2 arises when we choose b from two of the tree parentheses, therefore giving us

$$\binom{3}{2} = 3 \quad \text{of the } ab^2\text{-products.}$$

Finally, if we note that a^3 and b^3 arise with 0 or 3 choices of b, respectively, from the parentheses, then we can write

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$
$$= \binom{3}{0}a^3 + \binom{3}{1}a^2b + \binom{3}{2}ab^2 + \binom{3}{3}b^3$$

Expression (1) is of course the one we would use in practice but expression (2) helps us to see how the formula would look for higher powers. We can follow analogous choice logic for $(a + b)^5$, to take another example. It may, of course, also be written as

$$(a + b)(a + b)(a + b)(a + b)(a + b) \text{ and gives therefore}$$

which calculates out to

$$(a + b)^5 = \binom{5}{0}a^5 + \binom{5}{1}a^4b + \binom{5}{2}a^3b^2 + \binom{5}{3}a^2b^3 + \binom{5}{4}ab^4 + \binom{5}{5}b^5$$

which calculates out to

$$(a + b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5.$$

Power expansions of such binomials thus give formulas containing the numbers we encountered in Pascal's triangle. It is in connection with

the expansion of powers of binomials that these numbers have received the name "binomial coefficients."

3.2.5 *An Application in Physics (Theory of Heat)*

A solid cube of steel has sides of dimension 5 cm at 0°C. It is heated to 20°C, expanding so that each side has the length $5 + h$ cm, where h according to heat theory would be

$$h = 5 \cdot 20 \cdot 0,000012.$$

The decimal fraction is the so-called coefficient of expansion (for steel, in this case).

We get $h = 0,0012$.

What is this new volume?

We must expand $(5 + h)^3$ and get

$$(5 + h)^3 = 5^3 + 3 \cdot 5^2h + 3 \cdot 5 \cdot h^2 + h^3 = 125 + 75h + 15h^2 + h^3.$$

Of these terms, 125 is the volume before expansion (5^3). 75h has the value $75 \cdot 0,0012 = 0,09$ and is thus quite small. We can ignore the still smaller terms $15h^2$ and h^3. The new volume has, with good accuracy, the value $125 + 75h = 125,1$ cm³.

3.2.6 *The Surprising Triangle*

We have seen that Pascal's triangle is made up of the binomial coefficients (3.2.3 and 3.2.4). We have further seen (in 3.2.2) that the Greek triangle numbers form a line in the triangle. Are there other interesting number sequences which appear in Pascal's triangle? There are *many*, and it is often surprising that the triangle in one way or another contains a number sequence which one has come upon in some particular problem.

Let me give two examples to start with.

Example 1: If we add the numbers in Pascal's triangle row by row we obtain the sums:

1
1 + 1 = 2
1 + 2 + 1 = 4
1 + 3 + 3 + 1 = 8
etc.

The sums thus form the doubling sequence 1, 2, 4, 8, ... Will this continue with each succeeding row? (Exercise 2).

Example 2: In Moser's circle subdivision problem (Section 3.1.6) we obtained the following number of subdivisions:

1, 2, 4, 8, 16, 31, 57, 99, ...

We recall there that the series could fool us when it deviates from the doubling series 1, 2, 4, ... starting with the value 31. Can Moser's sequence, too, be found in Pascal's triangle? In fact, yes. If we add along the rows up to the line drawn in Figure 3.2.5 we get the correct sums. One can show that this applies as far as one wishes to go in the circle subdivision problem, or in Pascal's triangle.

	Maximum number of areas
1	1
1 1	2
1 2 1	4
1 3 3 1	8
1 4 6 4 1	16
1 5 10 10 5 / 1	31
1 6 15 20 15 / 6 1	57
1 7 21 35 35 / 21 7 1	99
1 8 28 56 70 / 56 28 8 1	163

Figure 3.2.5

3.2.7 A Glimpse of Probability, Chance and Risk

In the 9th grade the students are at an age when they particularly want to test their powers of intelligence, especially in discussions with

teachers or with parents at home. It is an important stage in their freeing themselves from dependence on adults. In mathematics class we get into the concept of probability; we pose problems concerning chance and risk. Some students have already met up with questions of the type: What is the chance of guessing all right on the football pools? Is it equally difficult to guess all of them wrong as all of them right? What is the chance of throwing three sixes in a row with a die? And so on. These are problems which they gladly investigate, so that they can feel they have clear-cut answers and that they have a grasp of the basic "foundation." I emphasize "foundation" here because the concept of probability is difficult, a difficulty which it is not easy to become conscious of. On the other hand, students readily note that the actual problems are themselves quite difficult, in fact "sneaky." One can easily be led astray without knowing it.

The students will once again meet Blaise Pascal, the man who together with his countryman Pierre Fermat laid the foundations of classical probability theory, and they will have the chance to tackle basic problems of the same type as Pascal faced.

Let us examine three problems which are related to earlier sections in this chapter.

1. Suppose we guess the answer to each of 5 questions which are to be answered with a "yes" or a "no." The questions are such that we do not have the slightest idea of the right answers. The chance of answering correctly is thus ½.

a) What is the probability of getting 3 correct answers?
b) What are the chances of getting at least 3 correct answers?

We start with a): it does not matter *which* 3 questions we succeed in guessing correctly. We could answer correctly on the first three, on the last three, on questions 1, 3, and 4, etc. How many such combinations of 3 right answers are there?

The number must obviously be equal to the number of ways of choosing 3 questions from out of 5, that is $\binom{5}{3} = 10$. (See 3.2.1-3.2.2).

There are thus 10 so-called successful cases, 10 different 5-row betting pool guesses with exactly 3 correct answers. But how many different 5-row bets are *possible at all*? On every row (question) we have two alternative answers. How many alternative answers does that give for 5 questions? Many pupils have a tendency here to answer $5 \cdot 2 = 10$ alternatives. This answer cannot be right, of course, since the number of alternatives with only three correct answers is 10.

In actual fact the number of alternative answers increases *multiplicatively* from row to row: for *each* of the two answers to question 1 we have two possible ways of answering question 2, which gives $2 \cdot 2$ alternative ways of answering both questions, etc. Figure 3.2.6 ought to show clearly enough why the growth is multiplicative. The number of *possible* alternative answers is therefore

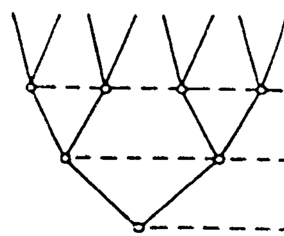


Figure 3.2.6

$$2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32$$

According to classical probability theory, the probability of exactly 3 correct answers is the following fraction:

$$\frac{\text{number of successful cases (with 3 correct answers)}}{\text{number of possible cases (with 0, 1, 2, 3, 4 or 5 correct answers)}}$$

The probability in question is thus

$$\frac{10}{32} \approx 0.31 \quad \text{or } 31\%$$

b) Getting at least 3 right means getting 3, 4, or 5 right. The probability sought after is therefore (according to the investigation above)

$$\frac{\binom{5}{3} + \binom{5}{4} + \binom{5}{5}}{32} = \frac{10 + 5 + 1}{32} = \frac{1}{2} = 50\%$$

(Can examples of this type of problem be solved yet more easily? See example 3 below!)

2. We have a combination lock, outfitted with 10 buttons labeled with the numbers 1, 2, 3, ... 9 and 0, plus a release button marked k. The lock opens only if one pushes down the right combination of number buttons and thereafter pushes k.

It does not matter in which order one pushes the number buttons. How big is the risk that an unauthorized person might succeed in opening the lock on his first try?

What do we mean by "risk"?

We take the same approach as Fermat and Pascal and seek the number of ways of setting the right combination in relation to the number of ways of pushing down *any* combination of buttons.

Let us even include the possibility of pushing k directly, i.e. the possibility of not setting any of the number buttons at all, as one of the possible ways.

The number of button combinations which opens the lock is obviously only one.

How many are then the total number of settings? A strategy which lies close at hand would be to calculate

the number of ways to choose no button $= 1$
plus the number of ways to choose 1 button $= 10$
plus the number of ways to choose 2 buttons $= \binom{10}{2}$
plus......
plus the number of ways to choose 10 buttons $= \binom{10}{10}$

This sum is equal to the total number of button combinations possible. But there is a considerably easier way to go about it.

We can look at the choice of buttons to push from another angle: for every button we suppose we face a choice shall we push the button or not? We have these two alternatives for each and every one of the 10 buttons.

Analogous to the study in Example 1, this gives $2^{10} = 1024$ possible combinations. The risk that the lock is opened after only one random try is thus

$$\frac{1}{1024} \qquad \text{or approximately } 0{,}1\ \%.$$

3. What are the chances of getting at least 1 six in 4 throws of a die? "At least 1 six" means 1, 2, 3, or 4 sixes. We might choose here the method of calculating

the number of cases with 1 six
the number of cases with 2 sixes
the number of cases with 3 sixes
and the number of cases with 4 sixes (which is 1)

and thereafter total up these numbers. If the total is x, then the chance we are looking for would be

$$\frac{x}{6 \cdot 6 \cdot 6 \cdot 6} = \frac{x}{1296}$$

since the number of possibilities is 6 at every throw.

There is, however, a time-saving trick, namely that of calculating the number of cases which are *opposite* of getting at least 1 six. This means simply determining the number of cases where no six appears at all during the four throws.

There are 5 such possibilities at *each and every* throw, which means the number of possible series of four throws without sixes is

$$5^4 = 625.$$

The number of cases with at least one six must then be

$$1296 - 625 = 671$$

and the possibility of throwing such a case would thus be $\frac{671}{1296}$

which gives a 51.8 % chance.

3.2.8 *Exercises*

1. Figure 3.2.2 showed the 10 paths between corners AO, BO and A2, B3. According to what principle are these paths ordered?

2. Prove that the sums in Example 1, Section 3.2.6, always give powers of 2, that is, that they give the sequence 1, 2, 4, 8, 16, ... (otherwise written $2^0, 2^1, 2^2, 2^3, 2^4, \ldots$) by setting x = 1 in the expansion of $(1 + x)^n$ for n = 0, 1, 2, 3, ...

3. In how many ways can 7 people be placed in a row, for example, along one side of a long table?

4. How many committees of four members can one select from seven people?

5. In Sections 3.2.2 and 3.2.3 we had examples of equalities such as

$$\binom{12}{5} = \binom{12}{7} \quad \text{and} \quad \binom{17}{11} = \binom{17}{6}.$$

Our motivation for this implies quite generally that

$$\binom{n}{k} = \binom{n}{n-k}$$

This relation can also be proven with the aid of the formula

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

How can this be shown?

6. On a circle lie 10 points. How many chords (connecting lines) can be drawn between these points?

7. 12 points are given. They lie such that no straight line can be drawn through any set of 3 points. How many triangles can be formed with three of the 12 points as corners?

8. How many letter groupings of 5 letters can be formed of the five letters A, B, C, D and E

   a) when each letter may only appear once in any grouping?
   b) when the condition above is relaxed, i.e. repetition of letters is allowed?

9. We return to Problem 2, section 3.2.7. As we remember, a lock has 10 buttons, labeled 0, 1, 2, ... and 9 respectively. If certain buttons are pressed, perhaps none at all, then the lock may be opened by pushing the release button. It does not matter in which order the proper buttons are pushed down.

How big is the risk that a thief opens the lock on one of 200 r[andom] tries (among which he may repeat some earlier trial combination)? Example 3 in Section 3.2.7 gives a clue.

3.3 Fibonacci Numbers

3.3.1 Fibonacci

Fibonacci ("son of Bonacci"), or Leonardo of Pisa as he wa[s] called, was one of those who most strongly contributed to the penetra[-]tion into Western Europe of Indian numerals, thereby pushing out th[e] Roman numerals then in use. Fibonacci grew up in Bugia on the North African coast along the Mediterranean. (The town is called Bejaij[a] today.) His father, Guglielmo Bonaccio, had a job which most closely could be called manager of a department store, which he owned. He was influential and wealthy. His son received his basic education from a Moorish teacher. He became acquainted with the Indian numbers, which had been assimilated into the Arabian culture. As a young man the son had the great advantage of visiting different countries in the Mediterranean area. He travelled in Egypt, Syria, Greece, Sicily, and Southern France and concerned himself with the arithmetical methods in use, including usage in commerce. Fibonacci found the Indian-Arabian system far superior to other systems and decided to write a book on the art of arithmetic.

Fibonacci's book *Liber Abaci* (abacus book; abacus = counting board) came out in 1202. It includes 15 chapters covering different areas of arithmetic. The first chapter is devoted to Indian numerals. Fibonacci thereafter takes up arithmetic with whole numbers, with fractions, business arithmetic, roots and square roots, equations, algebra and solution of selected problems. *Liber Abaci* became a great success and came out in a revised new edition in 1228. Beyond this volume, Fibonacci wrote three others, including one on geometry and disputed two of them in the presence of the benign Emperor, Fredrik II. Mathematics did not stand high at the universities of that time, however, and preference for the Roman

numerals lived on. Bankers were terrified with the thought that Indian numerals would invite easily-made falsification, for example, with extra zeroes or by erasing zeroes at the end of a number. It was not until the 1400's that bankers succumbed, but today we still have the custom of writing the amount with letters as well as numerals when we write a check.

Yet Fibonacci has become a name in the mathematical world not so much for having promulgated the use of Indian numerals and having shown their advantages, as for having posed an interesting problem. It is called Fibonacci's Rabbit Problem.

3.3.2 "Fibonacci's Rabbit Problem"

This problem was introduced in the second edition of *Liber Abaci*. We will now concern ourselves with the rabbit problem which, although it must be said to be rather artificial, nonetheless was shown great interest.

Our starting point is a pair of rabbits, one of each sex. Let us say this pair is newly born in month 0. In month 1 the couple is sexually mature and in month 2 they beget a pair of bunnies, also one of each sex. The original pair goes on giving birth to a new pair each month (and by "pair" we always mean one of each sex). Each newborn pair increases the family in the same manner as the original pair, i.e. beginning in month 2 of their lives they give birth to a new pair each month thereafter. We suppose further that none of the rabbits die.

We now ask: how many rabbit pairs are there after a year, i.e. in month 12? We might perhaps formulate the problem a little more specifically: how does the number of pairs grow month after month up to and including month 12? The task is then to construct a table of the number of rabbit pairs as a function of the month number. If we call the number of pairs A and add a subscript with the number of the month, we can then read the following directly from the Fibonacci text:

$$A_0 = 1 \quad A_1 = 1 \quad A_2 = 1 + 1, \quad \ldots, \quad A_{12} = ?$$

I have posed this problem many times in the tenth grade: "Try to keep track of the rabbits by making an illustrative figure, perhaps like a table, perhaps with different symbols for newborn, etc. Perhaps someone wishes to draw a sort of tree of the rabbit population. Try now to find a good systematic method for organizing and accounting for the rabbit population."

Since the problem's solution does not require any real mathematical knowledge, it is not surprising that everyone in the class considers the problem, starts with some idea or other, and begins to count rabbits. After a while the students begin comparing some of their preliminary results with each other, and groups begin to form. Some students begin perhaps with one method but find it unpromising (usually the work lacks clarity) and look for a new idea. In some classes there would be one student — usually one who had difficulty with the subject — who succeeded in coming up with a clear and correct solution while highly talented students got entangled in their diagrams.

Apart from the degree of success, it was evident that Fibonacci could engage the students' interest, and most of them gained experience of value for problem-solving in general.

We will now consider a few variants of rabbit counting which students presented to each other.

1. Trying to draw a sort of population tree did not work out well. With time, as the months progress, the tree had so many branches and shoots in different directions that the picture was unclear.

2. A method drawing a separate family tree for each pair of rabbits used incredible amounts of paper. The student marked off the rabbit couples with X-es and carried on as in Figure 3.3.1 (but continued farther on). Time seemed too short for this method, but the idea was undeniably direct.

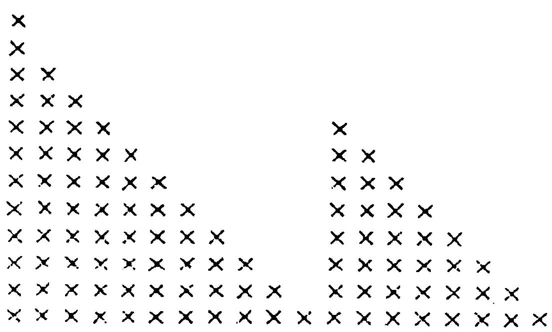


Figure 3.3.1

3a. Each pair was marked with a one, the birth of a new pair with a line connecting parents to children (Figure 3.3.2). The number of ones in each column is added up and in this way we get the desired A-values.

3b. As in solution 3a. each pair is marked with a 1 in the beginning but here each pair is only represented once with a 1, namely in the month they are born. When later several pairs are born in the same month, for example 5 pairs, then 5 ones are noted in the table as the increase for that month. The numbers which are written into the table are thus always newly born pairs. To get the total number of pairs at a given time, e.g. month 7, we must add all the ones up to and including month 7. See Figure 3.3.3.

4a. Circles represent newborn pairs; circles with dots are mature pairs; and darkened circles are pairs which are at least 2 months old ("aging pairs"). Month by month the aging, mature and newborn pairs are filled in (Figure 3.3.4).

4b. The same method as in 4a. but the letters A, M and N are used instead of the circle symbols for Aging, Mature and Newborn respectively (Figure 3.3.5).

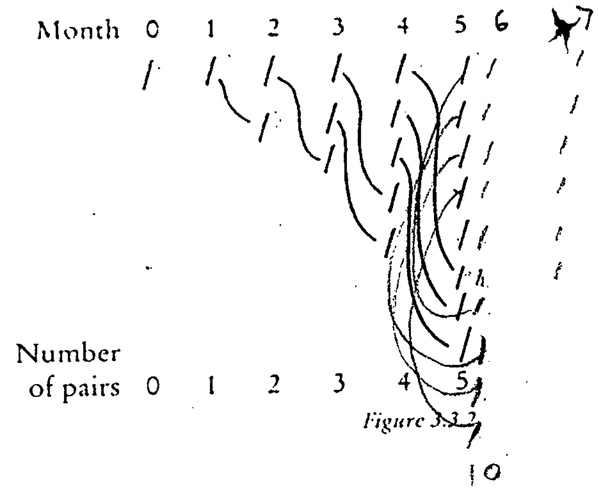


Figure 3.3.2

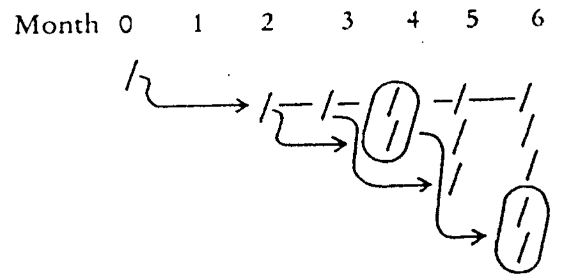


Figure 3.3.3

Month		Number of pairs
0	○	1
1	⊙	1
2	● ○	2
3	● ⊙ ○	3
4	● ● ⊙ ○ ○	5

Figure 3.3.4

The question now lies close to hand: how many surviving (aging and mature) and newborn shall we note down in the step going from one month to the next? Empirically we may have found some values, for example

Month		Number of pairs
0	1N	1
1	1M	1
2	1A + 1N	2
3	1A + 1M + 1N	3
4	2A + 1M + 2N	5

Figure 3.3.

$$A_3 = 3 \qquad A_4 = 5 \qquad A_5 = 8$$

By this time many pupils, alone or in groups, have already observed a regularity in the numbers obtained. Let us look at these:

$$A_0 = 1 \qquad A_4 = 5$$
$$A_1 = 1 \qquad A_5 = 8$$
$$A_2 = 2 \qquad A_6 = 13$$
$$A_3 = 3 \qquad A_7 = 21$$

These are enough values. Do we see any relation between the numbers? Some pupils, who have discovered a relation, choose to use it "mechanically" in continuing on to month 12, convinced that the regularity always holds.

We discover that each new number is the sum of the two nearest preceding numbers: 1 + 1 gives A_2, 1 + 2 gives A_3, 2 + 3 gives A_4, etc. Finally we have

$$A_{12} = A_{10} + A_{11} = 89 + 144 = 233.$$

But does this addition rule apply unconditionally? Might the answer to his question perhaps lead to a better method of doing the actual rabbit accounting?

5. If not earlier, then now some of the pupils will come upon the idea of asking: how many rabbit pairs live on, and how many are newborn when we go over from one month to the next?

If we consider, for example, the change from month 4 to month 5, then we have:

(1) All pairs of rabbits from month 4 live on, i.e. $A_4 = 5$ pairs

(2) In month 5 as many new are born as there were mature and aging pairs in month 4. This is as many as the total number of pairs in month 3, i.e. $A_3 = 3$ pairs. (See the table in Figure 3.3.4 or Figure 3.3.5.)

From this we get $A_5 = A_4 + A_3 = 5 + 3 = 8$.

The same reasoning gives $A_6 = A_5 + A_4$ and generally

$$A_n = A_{n-1} + A_{n-2} \quad (n = 2, 3, 4, \ldots) \tag{1}$$

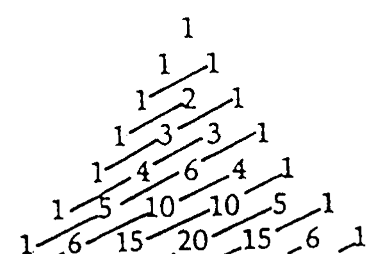
With this newly proven formula we can successively obtain the number of rabbit pairs in any month we wish, but we must use the formula repeatedly step by step.

Is there not perhaps some formula for *direct* calculation of the number of pairs "so that we can get out of adding up all the lower numbers"? This question lies near at hand to the pupils. Since this problem is quite tricky and would require considerable time to develop, I usually content myself with a formula which the French mathematician J. P.M. Binet published in 1843. In order to keep the formula as simple as possible we renumber the Fibonacci numbers (1,1, 2, 3, 5, 8, ...) with f_1 as the first number of A: $f_1 = 1$, $f_2 = 1$, $f_3 = 2$, $f_4 = 3$, etc.

The formula then is

$$f_n = \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right\}, \; n = 1, 2, 3, \ldots$$

This formula was actually known in the 1730's to the Swiss mathematicians Leonard Euler and Daniel Bernoulli. There is, however, from the standpoint of the pupils, a simpler formula for direct calculation of lower Fibonacci numbers, proven by É. Lucas. Like so many other number relationships it may be found hidden in Pascal's triangle. In Figure 3.3.6 parallel



Fibonacci numbers in Pascal's triangle diagonals with a specific slope are drawn. If we sum the numbers which lie on such a diagonal, beginning from the top, we will obtain precisely the Fibonacci numbers. An appropriate and stimulating task is to prove, with the help of formula (1), that all of the diagonals — not only those in the beginning — give Fibonacci numbers.

We have here a convenient addition formula for the Fibonacci numbers:

$$f_1 = 1 = \binom{1}{0}$$

$$f_2 = 1 = \binom{1}{1}$$

$$f_3 = 2 = 1 + 1 = \binom{2}{0} + \binom{2}{1}$$

$$f_4 = 3 = 1 + 2 = \binom{3}{0} + \binom{2}{1} \text{ etc.}$$

and generally $\quad f_n = \binom{n-1}{0} + \binom{n-2}{1} + \binom{n-3}{2} + \ldots \qquad \text{for } n \geq 1,$

where the summation extends until the difference between the upper and the lower numbers is zero in the case of n being odd, or one, in the case of even n.

It may now be of value to review the different rabbit-counting methods developed by pupils which were presented as variants 1 - 5 above. We can differentiate various degrees of abstraction:

In 3a, 3b and 4a we count up ones or figures which represent rabbit pairs, more or less as we would count several litters of animals. The degree of concreteness is high here; the achievement is almost entirely in the construction of a systematic arrangement of the units.

In 4b, letters are introduced in a first step toward abstraction, as symbols for the different kinds of rabbit pairs.

In method 5 we leave both the symbols and the tangible counting of rabbit pairs and find the additive structure of the Fibonacci numbers. We have here reached the level of thought which gives us a general method, even if it is a step-by-step method. (Formula (1) is called a recursive formula.)

The more levels of concreteness and abstraction with which a particular problem may be solved, the more suitable the problem is for differentiation within a class.

3.3.3 *More on Fibonacci Numbers*

Why have Fibonacci numbers become so well known? What interesting characteristics do they show? It became apparent that Fibonacci numbers more-or-less unexpectedly appeared in many different contexts and that such numbers have a whole variety of relationships between themselves.

Apart from certain aspects which are taken up in the exercises below, we will now primarily concentrate on the occurrence of Fibonacci numbers in the plant kingdom and show their relation to the so-called "golden section."

Let us first return to our rabbit population and ask: how large are the fractions of newborn pairs and surviving pairs, respectively, month by month? Is it a population which grows younger and younger in the sense that the fraction of newborn increases, or is it an aging population?

We select the following notation:

A_n = number of rabbit pairs in month n (n = 0, 1, 2, ...)
B_n = number of newborn pairs in month n.

The fraction of newborn is then $u_n = \frac{B_n}{A_n}$. The fraction of others (surviving) is $v_n = 1 - u_n$.

Taking n up through 12 we get the following table:

n	u_n	v_n	n	u_n	v_n
0	$\frac{1}{1}=1$	$\frac{0}{1}=0$	7	$\frac{8}{21}=0{,}38695\ldots$	$\frac{13}{21}=0{,}61904\ldots$
1	$\frac{0}{1}=0$	$\frac{1}{1}=1$	8	$\frac{13}{34}=0{,}38235\ldots$	$\frac{21}{34}=0{,}61764\ldots$
2	$\frac{1}{2}=0{,}5$	$\frac{1}{2}=0{,}5$	9	$\frac{21}{55}=0{,}38181\ldots$	$\frac{34}{55}=0{,}61818\ldots$
3	$\frac{1}{3}=0{,}33\ldots$	$\frac{2}{3}=0{,}66\ldots$	10	$\frac{34}{89}=0{,}38202\ldots$	$\frac{55}{89}=0{,}61797\ldots$
4	$\frac{2}{5}=0{,}4$	$\frac{3}{5}=0{,}6$	11	$\frac{55}{144}=0{,}38194\ldots$	$\frac{89}{144}=0{,}61805\ldots$
5	$\frac{3}{8}=0{,}375$	$\frac{5}{8}=0{,}625$	12	$\frac{89}{233}=0{,}38197\ldots$	$\frac{144}{233}=0{,}61802\ldots$
6	$\frac{5}{13}=0{,}38461\ldots$	$\frac{8}{13}=0{,}61538\ldots$			

Can we see anything interesting in this table?

Some students notice that the fractions seem to converge toward the values 0.3820 and 0.6180 respectively. Others notice that u_n and v_n swing up and down in value — every other time increasing, every other time decreasing — but the swings become smaller and smaller as n increases. With the aid of a calculator one can easily extend the table for a few more values of n. One finds, for example,

$$A_{18} = 4181,\ A_{19} = 6765,\ A_{20} = 10946, \text{ of which}$$

$$u_{20} = \frac{4181}{10946} = 0.38196601$$

$$v_{20} = \frac{6765}{10946} = 0.61803398$$

In fact it does seem as if, decimal by decimal, the numbers are converging to certain values. Judging by everything, the fractions u_n and v_n are approaching limiting values

$$g = 0.38196601 \quad \text{and } G = 0.61803398\ldots$$

The fractions u_n and v_n seem to come as close as we wish to two numbers g and G respectively and their approaches are oscillatory. We let rest the questions of the existence and of the exact values of g and G until Exercises 4 and 5, and ask here instead: *if* limits exist, do they then eventually appear with the same values when one begins with entirely different starting values for the number of pairs in months 0 and 1? For example, if we have 6 newborn pairs and 1 mature pair in month 0? In this case we get a Fibonacci series which begins with

$$A_0 = 7,\quad A_1 = 8 \text{ (7 pairs + 1 newly born pair)}$$

and which continues with $A_2 = 15$, $A_3 = 23$, $A_4 = 38$, etc.

What happens to the functions u_n and v_n respectively? This problem is taken up in Exercises 3-6.

3.3.4 Leaf Rotation (Phyllotaxis)

Fibonacci numbers lay more-or-less in a "Sleeping Beauty" state for 600 years, until the 1830's. We have seen that Euler, D. Bernoulli, and Binet had studied them the century before, but they were given at least as much attention by the botanists of the 19th century.

Alexander Braun published in 1831 a dissertation with the title "Vergleichende Untersuchung über die Ordnung der Schuppen an den Tannenzapfen als Einleitung zur Untersuchung der Blattstellungen überhaupt" (Comparative study of scale arrangement on pine cones as introduction to a general study of leaf position).

As the title suggests Braun studied the geometry of pinecones. Anyone can understand from the appearance of a pinecone (Figures 3.3.7 and 3.3.8) that the cone has an architecture which one ought to describe mathematically. Braun's research lead to the concept of leaf fraction or phyllotaxis.

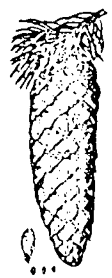
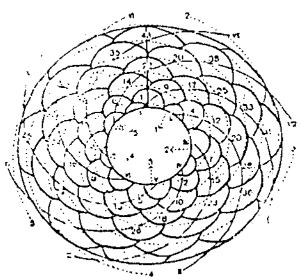


Figure 3.3.7
Pinecone



Pinecone with phyllotaxis 8/21
(From Strasburger, "Lehrbuch der Botanik")

Goethe had emphasized that ordinary upright, growing flowering plants with leaves follow two basic principles during growth: the stalk grows straight upward (the vertical principle) while the leaves (usually) form an upward spiral. As the leaves grow outward from higher and higher levels there occurs a "twisting" around the stalk from one leaf to the next highest neighbor. Besides the vertical principle there is thus also a "rotation" with the stalk as axle (Figure 3.3.9). When the plant flowers, the leaf spiral is completed with a crown of leaves (the petals). Braun showed that leaf rotation is by different amounts in different categories of plants. We shall here take a short look at some phyllotaxis results which are well known to botanists.

1. On tulips and gladiolas the leaves grow out alternately to the left and to the right. The "twisting" from one leaf to the next is thus 1/2 revolution. The phyllotaxis (leaf fraction) is said to be 1/2. (Figure 3.3.10).

2. In 3-bladed grasses and meadow saffron the three leaves distribute themselves around 1 revolution, so that the rotation from one leaf to the next is 1/3 revolution, i.e., the phyllotaxis is 1/3 (Figure 3.3.11).

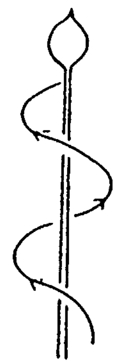


Figure 3.3.9

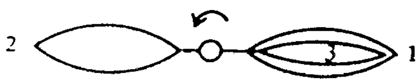


Figure 3.3.10

3. If we count the leaves in plants of the Rosacae family (roses, raspberries, plums, violets, etc.), we find that 5 successive leaves upwards on the stem twist 2 revolutions. The twist is thus 2/5 revolution per leaf and the phyllotaxis is 2/5 (Figure 3.3.12).

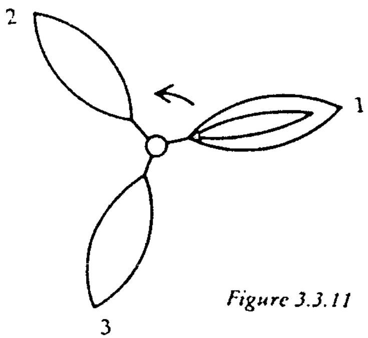


Figure 3.3.11

4. In most cabbages (the Cruciferae family), snap dragons, plantain and monk's hood there are 8 leaves to every 3 revolutions and the phyllotaxis is 3/8.

5. Dandelions, mullein (figwort family), potatoes and other plants have a phyllotaxis 5/13.

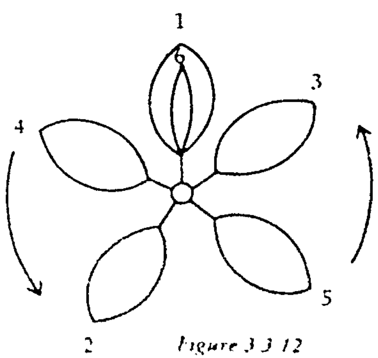


Figure 3.3.12

6. Now we come to the starting point of Braun's investigations: the fir. It has a phyllotaxis of 8/21, both in the scales on the cone and in the rotation of the needles. The same values apply to the pines and larches.

In summary: phyllotaxes of 1/2, 1/3, 2/5, 3/8, 5/13 and 8/21 have been found.

What numbers do we have in this sequence? Fibonacci's numbers appear, as we see, in the phyllotaxes, and the

fractions themselves agree with the population fractions which we computed in the previous section 3.3.3.

If phyllotaxes were to exist with even higher Fibonacci numbers, such fractions should approach the value g = 0.3820... Are there such fractions? Can it be, that the limit g has a very real meaning in the plant kingdom? These are questions which it is very natural to ask. The botanist G. van Iterson counted leaves up the stalk and used a microscope when the naked eye could no longer determine the point of origin of the very tiny stems on leaves which were just coming out. These studies, presented by Iterson in 1907 in the dissertation "Mathematische und mikroskopisch-anatomische Studien über Blattstellungen" and later furthered by M. Hirmer (1922), concentrated on the vegetative cone, that part of the plant which produces the very youngest leaf buds. The results showed that these start, independent of botanical family, with the leaf fraction g ≈ 0.38 and that the phyllotaxis then deviates from this more and more as we go down the stalk to older leaves. There the leaf rotation finally becomes what is typical for that family.

All leaf rotations thus begin with the value g ≈ 0.38 at the top leaf buds. The phyllotaxis thereafter gradually changes toward the family's characteristic fraction, for example 2/5 = 0.40 for the roses.

Further examples of Fibonacci numbers in the plant world are given in Section 3.7.2.

3.3.5 The Golden Section

The results from research on phyllotaxis becomes even more interesting if we relate them to a proportion which was much appreciated during the Middle Ages and known even before that by the Greeks, the proportion which arises from the so-called golden section. Let us take a straight line AB of length 10 cm, for example. A point S on AB divides the length into the golden section if the ration of the smaller interval to the larger is the same as the ration of the larger to the whole interval.

In other words, if AS is the shorter length, the golden section would imply

$$\frac{AS}{SB} = \frac{SB}{AB} \quad \text{or letting } x = SB\text{:} \quad \frac{1-x}{x} = \frac{x}{1}$$

In order to illustrate this sectioning we let the rectangle in Figure 3.3.13 have length 1 unit and the shorter side width x. At one end of the rectangle we mark off a square with side x. According to (1) the rectangle is now divided into golden section proportions if the smaller rectangle is similar (in the geometrical meaning) to the larger, original rectangle. If one solves equation (1), leading to the second degree equation

$$x^2 + x = 1,$$

one gets as the useful (positive) root the number $x = \frac{\sqrt{5} - 1}{2} \approx 0.6180\ldots$

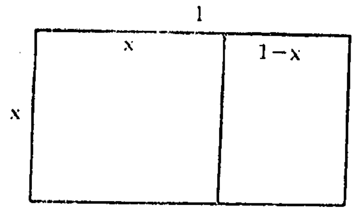


Figure 3.3.13

This number determines the golden sectioning of the interval AB and agrees exactly with the limit G which arose in the study of leaf rotation (Section 3.3.4). (This agreement is proven in the solution to exercises 4 and 5.)

When S is placed at the golden section's dividing point, the decimeter long interval is divided into two parts g and G, approximately 0.38 and 0.63 respectively. Figure 3.3.14 shows this sectioning.

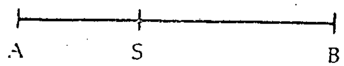


Figure 3.3.14

Let us now construct a triangle in which the base is to the side as G is to 1. We take the side AB = 1 dm (decimeter = 10 centimeters), for example, and the base BS = G dm ≈ 0.62 dm (Figure 3.3.15a). We measure the top angle with a protractor and find that it is 36°. Is this value exact? If that were the case the triangle would comprise one-tenth of a regular ten-sided polygon, since the 10-side polygon's inner triangles have a center angle of

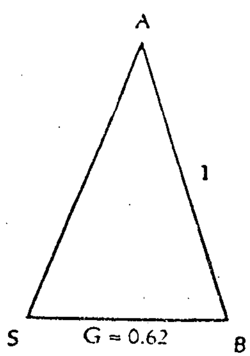


Figure 3.3.15a

$$\frac{360°}{10} = 36°.$$

Does the golden section produce such triangles?

Let us instead begin with a regular ten-sided polygon, remove one of its triangles, and set its side equal to 1 unit. We ask, how long is the

triangle's base? Those who know the basis of trigonometry can directly write the relation

$$\frac{\text{half the base}}{\text{the side}} = \sin 18°$$

from which

$$\begin{aligned} \text{base} &= 2 \cdot \text{side} \cdot \sin 18° \\ &= 2 \sin 18° \\ &\approx 0.618 \end{aligned}$$

We are now well convinced that the length of the base is exactly G, i.e. that our golden-section triangle fits into the ten-sided polygon. But how could one show this exactly?

How large are the angles in the 10-sided polygon's triangles? Obviously the base angles are 72°, since 36° + 2 · 72° gives the correct sum of 72°, since 36° + 2 · 72° gives the correct sum of angles, 180°. If we now bisect one of the base angles with a line, we get two similar triangles as in Figure 3.3.15b.

The equation for equal rations of base to side in the small triangle BCD and in the large triangle ABC becomes

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

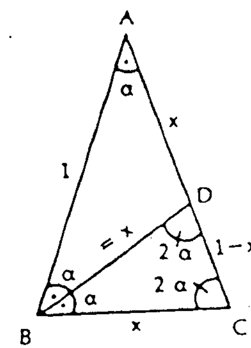


Figure 3.3.15b

This equation is the same as equation (1)! The useful root has already been given,

$$x = G = \frac{\sqrt{5} - 1}{2}$$

Ten golden-section triangles thus form a regular ten-sided polygon.

3.3.6 The Golden Section in Art and Nature

The golden section, as mentioned, was much appreciated during the Middle Ages and was used as the starting point in a variety of applications. A few examples:

The artist Piero della Francesca, who was both painter and mathematician, in all likelihood used geometrical structures as the starting point in a number of his paintings, and made use, in particular, of the golden section. Giotto in his famous painting of Franciscus with the birds placed the painting's center of interest , Franciscus' right eye, at the golden-section proportion along the painting's diagonal. A number of Gothic cathedrals show golden-section proportions, as do the shape of Stradivarius-violins.

In Section 3.3.4 we studied leaf rotations. We can formulate Iterson's and Hirmer's research results in the following way: in the budding leaves of plants with spiraling leaves there is a common leaf rotation equal to the golden section's smaller number, $g = 0.38$. From this value the various plants or families of plants then diverge toward the rotation fraction which is characteristic for them.

An interesting research report on proportions came out in 1854 in the dissertation by A. Zeising, "Neue Lehre von den Proportionen des menschlichen Körpers" (New Findings on the Proportions in the human body). As a result of various series of measurements, Zeising shows that on average the navel divides the human body height into the proportions of the golden section. In one class where the pupils became interested by this they measured and calculated their own values and found 0.617 as the mean value of the proportion foot-to-navel divided by body height (FN/BH).

Of course, there are those who dismiss Zeising's results as coincidence, but the golden section recurs repeatedly in the proportions of the arms and of the hands.

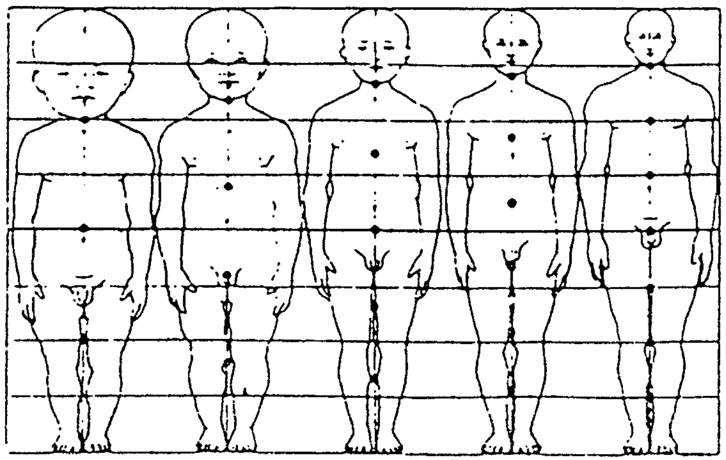


Figure 3.3.16

In the newborn child the proportion FN/BH is roughly 1:2. Figure 3.3.16 from Mörike-Mergenthaler's "Biologie des Menschen" ("Human Biology") illustrates how this proportion changes during growth. It is worth noting that the human being, in contrast to the leaves of the plants, *grows toward* the golden-section proportion.

3.3.7 Exercises

1. Drone bees develop from unfertilized eggs, in contrast to worker bees who grow from fertilized eggs. In other words, drones have only one parent, the mother, while workers and queen bees have two parents.

a) Draw a family tree for a drone bee, going 5 or 6 generations back. The figure here (Figure 3.3.17) shows the start of the tree.

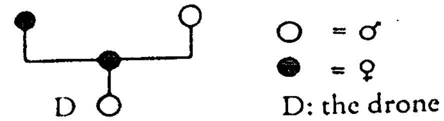


Figure 3.3.17

b) Do the numbers of bees in the tree, generation by generation, make up a Fibonacci series? If so, how can this be proven?

2. Fibonacci Puzzle

In Figure 3.3.18 is a square with 21 units as side (21r) divided into four parts – two right triangles A and B plus two trapezoids C and D. A and B together form a rectangle 8r x 21r. The shortest side in each of the four parts is 8r. In Figure 3.3.18b it appears as if the four pieces of the square form a rectangle with sides 13r and 34r. If this is the case, then the areas of the square and the rectangle should agree. But

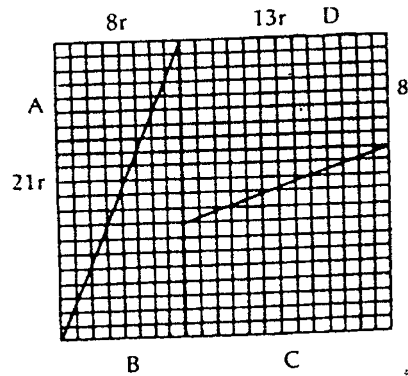


Figure 3

$(21r)^2 = 441r^2$ and $13r \cdot 34r = 442r^2$.

Where is "the catch" here?

3. Calculate to 3 decimal places, correctly rounded, the value of the quotient between the 9th and 10th numbers in the Fibonacci series which begins with the numbers 7 and 8. In other words, calculate f_9/f_{10} when $f_1 = 7$ and $f_2 = 8$. ($f_3 = f_2 + f_1$, $f_4 = f_3 + f_2$, etc.)

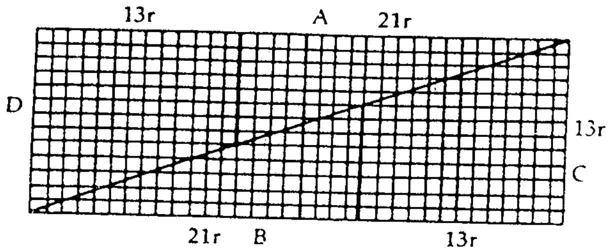


Figure 3.3.18b

4. Show that if the quotient of two successive numbers in the Fibonacci series 1, 1, 2, 3, 5, ... is

$$v_n = \frac{f_n}{f_{n+1}},$$

then the next quotient will be

$$v_{n+1} = \frac{f_{n+1}}{f_{n+2}} = \frac{1}{1+v_n}.$$

Show also, using this result, that if the quotients go toward a limit as n goes to infinity, then the limit is

$$G = \frac{\sqrt{5}-1}{2}$$

5. (For readers with experience in the convergence series)
How may it be proven that the number G (introduced in Sec. 3.3.4 and above in Exercise 4) truly is the limit of the sequence $\{v_n\}$ in Sec. 3.3.3.?

6. a) Form the 10th Fibonacci number if the first two are $F_1 = a$ and $F_2 = b$.

b) Does the value of the ratio between two successive Fibonacci numbers, F_n/F_{n+1}, approach a limit no matter what the two starting points are? Is this limit, if it exists, independent of the starting numbers, i.e. is there a common limit for all possible Fibonacci series? (Let a and b be positive.)

7. Show that the Fibonacci series which begins with G and 1 (that is the sequence G, 1, G + 1, 2G + 3, 3G + 5, ...) is identical with the geometrical sequence

$$G, 1, \frac{1}{G}, \frac{1}{G^2}, \ldots$$

8. For numbers in the Fibonacci sequence 1, 1, 2, 3, 5, ... it may be shown that

$$f_3 = 2 = 1 + 1 = 1^2 + 1^2$$
$$f_4 = 3 = 4 - 1 = 2^2 - 1^2$$
$$f_5 = 5 = 4 + 1 = 2^2 + 1^2$$
$$f_6 = 8 = 9 - 1 = 3^2 - 1^2$$
$$f_7 = 13 = 9 + 4 = 3^2 + 2^2$$

Does this regularity continue – and how can it be written as a general expression?

3.4 1, 2, 3, 4, . . . and 1, 2, 4, 8, 16, . . . Two Important Growth Principles

The sequences 1, 2, 3, 4, ... and 1, 2, 4, 8, 16, ... have become a part of everyday folklore through the classical story of the inventor of chess. According to the story, the inventor, who was promised whatever he wanted as a reward by an Indian prince, asked for one kernel of wheat on the first square of the chessboard, two on the second square and each time double the amount on each of the other 62 squares. The total number of kernels would be

$$1 + 2 + 4 + 8 + \ldots + 2^{63}$$

(where the last term is the product of 63 two's multiplied together).

Calculated out this became some 18 billion billion kernels of wheat, or more than the harvest in the whole province!

If the inventor had asked for 1 kernel on the first square, 2 on the second, 3 on the third, etc., then the reward would have remained at the modest level of

$$1 + 2 + 3 + 4 + \ldots 64 = \frac{64 \cdot 65}{2} = 2080$$

We shall soon see that both sequences of numbers in the title play important roles in various topical applications.

3.4.1 *The Piano Keyboard*

Figure 3.4.1 shows the keyboard of a piano, which most often spans 7 octaves plus 2 extra white keys and 1 extra black on the left.

Each octave has 7 white and 5 black keys which together give 12 tones. The so-called octave "just to the left of the piano's keyhole" goes from middle C to C in the octave above (c^1). If we count the white keys with C as the first, then we come to c^1 as the eighth key (and from this the name octave). c^1 is the first note in the next higher octave. Piano tuners tension the a^1 string so that it vibrates at 440 vibrations per second. The tone then has the frequency, as we call it, of 440 cycles per second, or 440 Hertz (Hz). Going from note a^1 to a^2 in the next octave, that

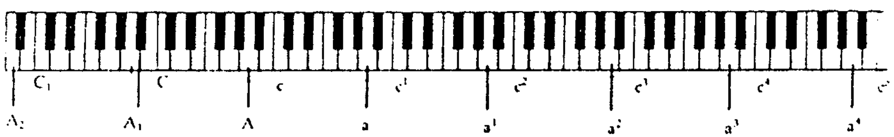


Figure 3.4.1

is taking a 1-octave step upward from a^1, we get a tone which, when played at the same time as a^1, is in perfect harmony with a^1. Playing the white keys from a^1 up to a^2, it is for our tonal experience as if returning to the starting tone, but on a higher level. We have taken a step upward on the tonal scale which feels completely natural.

The tonal ear is very sensitive to the exact agreement of a^2 relative to a^1. When tuned clearly, a^2 has exactly double the frequency of a^1, that is $2 \cdot 440 = 880$ Hertz.

In this way the frequencies increase octave by octave, and we have (in Hertz)

$$a^1 = 440$$
$$a^2 = 2 \cdot 440 = 880$$
$$a^3 = 4 \cdot 440 = 1760$$
$$a^4 = 8 \cdot 440 = 3520 \text{ Hz}$$

Taking a 1-octave step downward instead, i.e. to the left on the keyboard, we get (using the common nomenclature)

$$a = \frac{1}{2} \cdot 440 = 220 \text{ Hz}$$
$$A = \frac{1}{4} \cdot 440 = 110$$
$$A_1 = \frac{1}{8} \cdot 440 = 55$$
$$A_2 = \frac{1}{16} \cdot 440 = 27.5$$

The a-notes thus have the following frequencies:

$$A_2 = \frac{1}{16} \cdot 440 \qquad a^1 = 1 \cdot 440$$
$$A_1 = \frac{1}{8} \cdot 440 \qquad a^2 - 2 \cdot 440$$
$$A = \frac{1}{4} \cdot 440 \qquad a^3 = 4 \cdot 440$$
$$a = \frac{1}{2} \cdot 440 \quad \text{and} \quad a^4 = 8 \cdot 440$$

If we start with c instead of with a, the sequence will be similar, but 440 is replaced by another frequency. Let us call it f:

$$C_1 = \frac{1}{8} \cdot f \qquad c^2 = 2 \cdot f$$
$$C = \frac{1}{4} \cdot f \qquad c^3 = 4 \cdot f$$
$$c = \frac{1}{2} \cdot f \qquad c^4 = 8 \cdot f$$
$$c^1 = 1 \cdot f \quad \text{and} \quad c^5 = 16 \cdot f \quad \text{(which is the highest tone on our piano)}$$

The coefficients 1, 2, 4, ... always appear when we go upward by octaves and the coefficients ½, ¼, ⅛, ... when we go downward by octaves. In both cases 1 corresponds to the frequency we start with.

We now leave the world of tones temporarily and take a look at bacteria and radioactive substances.

3.4.2 *Bacteria and Radioactive Substances*

If a bacterial culture or a culture of yeast cells is allowed to grow under constant and favorable conditions, including sufficient nourishment available the entire time, then the number of bacteria or cells doubles with certain intervals of time. If the number of bacteria at time 0 is A, then this number will grow in the given time interval, let us say 3 hours here, according to the following table:

Time, hours	0	3	6	9	12	15
Number of bacteria	A	2A	4A	8A	16A	32A

But as soon as the nourishing culture begins to dry up, this regularity stops; growth is slowed. Growth can stop; the number of bacteria can even decrease. Figure 3.4.2 illustrates the table in diagram form.

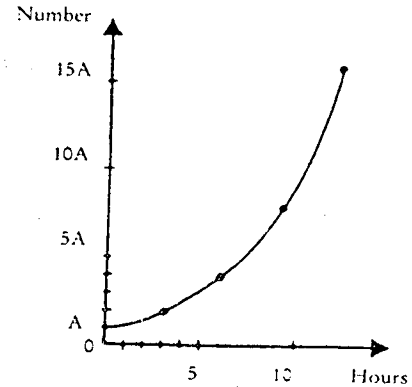


Figure 3.4.2

In yet another case do the doubling numbers 1, 2, 4, 8, ... play an important role.

Not long after the first radioactive elements had been discovered by the husband and wife Curie team (polonium in 1898 and radium in 1902), it was found that radioactive substances disintegrate and that during a certain specific time, for example 5 days, a substance would be reduced to only half of its original amount.

Whenever we begin our measurement, after the 5 days one half of the original amount of the substance will be left. Each radioactive substance has its own so-called half-life.

Uranium (with atomic weight 238) has a half-life of 4.5 billion years, radium (atomic weight 226) has 1600 years, polonium (atomic weight 210) 139 days, and a polonium isotope (atomic weight 214) just 0.00015 seconds.

If the quantity of material weighs m grams to start with and the half-life is T days, we have the following table illustrating the disintegrating principle:

Time (number of days)	0	T	2T	3T	4T	5T
Weight of substance in grams	m	½ m	¼ m	⅛ m	$\frac{1}{16}$ m	$\frac{1}{32}$ m

Here we see the halving numbers ½, ¼, ⅛, etc.

A curve of the decreasing substance weight is shown in Figure 3.4.3.

Earlier in Section 3.1 we wrote the numbers 1, 2, 4, 8, ... as powers of 2:

$$1 = 2^0$$
$$2 = 2^1$$
$$4 = 2^2$$
$$8 = 2^3$$
$$16 = 2^4 \quad \text{etc.}$$

If we go from the bottom up in this column, we find that the numbers (on the left) are halved each step upward while the exponents (the powers on the right) decrease by 1 with every step. For this reason 0 has been defined as the exponent of 2 for the number 1, i.e. $2^0 = 1$.

If we continue to halve on the left, we will get values

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \text{ etc.}$$

Can we also write these as powers of 2? The natural thing would then be to write

$$\frac{1}{2} = 2^{-1} \quad \frac{1}{4} = 2^{-2} \quad \frac{1}{8} = 2^{-3} \quad \frac{1}{16} = 2^{-4} \quad \text{etc.} \qquad (*)$$

We now draw up a table which shows how the inverse parts 2 and 1/2, 4 and 1/4, etc. also correspond to each other with regard to their exponents.

Why is the power form (*) so logical and natural?

The answer is quite simply that the rules which are used for arithmetic with positive exponents will also work with 0 and negative integers as exponents.

We content ourselves here with a few examples which show the applicability of the three main rules for exponent arithmetic.

Number	Power	Exponent
16	2^4	4
8	2^3	3
4	2^2	2
2	2^1	1
1	2^0	0
$\frac{1}{2}$	2^{-1}	−1
$\frac{1}{4}$	2^{-2}	−2
$\frac{1}{8}$	2^{-3}	−3
$\frac{1}{16}$	2^{-4}	−4

(1) $2^3 \cdot 2^4 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2$
$= 2^7 = 2^{3+4}$

(2) $\frac{2^3}{2^4} = \frac{2 \cdot 2 \cdot 2}{2 \cdot 2 \cdot 2 \cdot 2} = 2^{-1} = 2^{3+(-4)}$

(3) $(2^3)^4 = 2^3 \cdot 2^3 \cdot 2^3 \cdot 2^3 = 2^{3+3+3+3}$
$= 2^{3 \cdot 4}$

Rules: (1) $2^p \cdot 2^q = 2^{p+q}$

(2) $\frac{2^p}{2^q} = 2^{p-q}$

(3) $(2^p)^q = 2^{pq}$

These rules can be shown to apply for *any* choices of integers for p and q. In a number system such as this the exponents are called logarithms, or more precisely, base-2 logarithms, when the base is 2, as here.

One writes

$$\log_2 8 = 3 \quad \log_2 4 = 2 \quad \log_2 1 = 0 \quad \log_2 \frac{1}{2} = -1 \quad \text{etc.}$$

Analogously we may construct a table of the basic numbers in the 10-system and introduce base-10 logarithms:

$$\log_{10} 1000 = 3, \quad \log_{10} 100 = 2, \quad \log_{10} 10 = 1, \quad \log_{10} 1 = 0, \quad \log_{10} 0,1 = -1, \text{ etc.}$$

Can we also write numbers such as 17, 145, 5 and 2, 8 in power form? Can we determine (define) numbers x and y such that, for example,

$$13 = 10^x \text{ and } 13 = 2^y$$

in the 10- and 2-systems respectively?

Yes, we can and it has been done. The x and y values in question are called the base-10 and base-2 logarithms, respectively, of the number 13.

For every positive number whatsoever one can determine a base-10 logarithm or a base-2 logarithm or a logarithm in any other base so long as

it is positive. We cannot go into how one does this computation, but we can take a look at the graphical equivalent (more or less accurate depending on how well we draw the graphs) in the curves $y = 10^x$ and $y = 2^x$ respectively and the figures show how one graphically finds the logarithms $\log_{10} 13$ and $\log_2 13$.

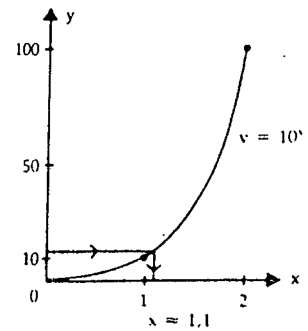


Figure 3.4.4
Graphical determination of $\log_{10} 13$

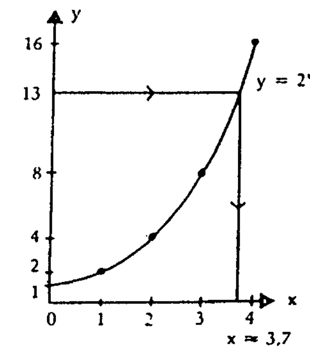


Figure 3.4.5
Graphical determination of $\log_2 13$

If we construct a scale for the numbers in the 10-system alongside a linear scale (equal spacing) for their base-10 logarithms, as in Figure 3.4.6, we call the resulting scale for the numbers logarithmic.

On the slide rule's main scales (usually called C and D and sometimes marked with an x) the numbers go from 1 to 10 and are fitted to a linear scale of logarithms going from 0 to 1 (Figure 3.4.7).[1]

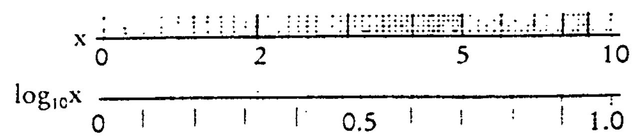


Figure 3.4.6
Logarithmic x-scale

We now return to the world of music and shall get to know something about the difficult task a piano tuner faces.

[1] Nowadays the slide rule is of interest only in connection with the conception of the logarithm and the pupil's understanding of it.

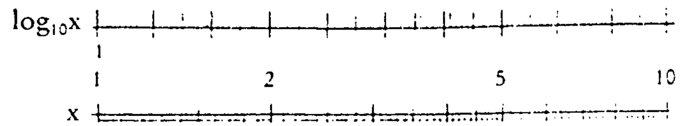


Figure 3.4.7
Slide rule scales

3.4.3 *The Piano Tuner's Difficult Task*

We know that the musical ear is very sensitive to octaves. The piano tuner must therefore tune pure octave steps; this means physically that the frequency exactly doubles from one tone to the same tone an octave higher.

A violinist, by contrast, tunes his of her four strings in fifths. The "fifth" interval corresponds to 7 so-called half-tones. (See Figure 3.4.8 for an example.)

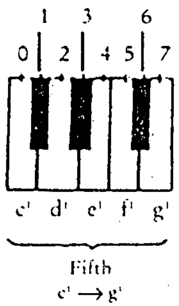


Figure 3.4.8

Counting from c^1, the fifth white key to the right is g^1 which is the next higher fifth.

The violin strings are tuned so that they give g, d^1, a^1, and e^2, forming pure fifth steps upward. A pure fifth step is, second to pure octave steps, that interval which the ear most easily distinguishes. In terms of frequencies the fifth step upward means that the frequency is multiplied by

$$\frac{3}{2} = 1.5.$$

Looking at a piano keyboard we can ask ourselves: how many fifths cover the 7 octaves from A_2 to a^4 or — which is the same thing — from C_1 to c^5?

We can count 7 half-tone steps at a time (equalling a fifth) and find as in Figure 3.4.9 that 7 octaves correspond to 12 fifths.

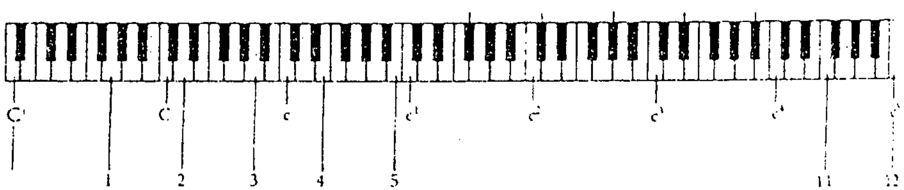


Figure 3.4.9

On the piano, 12 fifth steps (at 7 half-tones each) are equal to 7 octave jumps (at 12 half-tones each).

With A_2 as a starting point we have the frequency 27.5 Hz (cps). After 7 octave-sized steps, that is, after 7 doublings in frequency, we come to

$$a^4 = 128 \cdot 27.5 \text{ Hz}.$$

Let us instead take twelve fifth-sized steps from A_2. Here we get the frequency of a^4 to be

$$a^4 = 1,5 \cdot 1,5 \cdot 1,5 \cdot \ldots \cdot 1,5 \cdot 27.5 \text{ Hz or approximately}$$
$$a^4 \approx 129.75 \cdot 27.5 \text{ Hz}$$

But 129.75 · 27.5 must surely be bigger than 128 · 27.5! The twelve pure fifths give

$$a^4 \approx 129.75 \cdot 27.5 \approx 3568.1 \text{ Hz}$$

while the seven pure octaves give

$$a^4 = 128.27 \cdot 5 \approx 3520 \text{ Hz}.$$

How can this be reconciled? In no other way than that the piano tuner, among other things, makes all the fifths a little lower. His difficult task is in fact to tune the fifths (and other intervals) just impurely enough. When the degree of impurity is just right, the piano sounds good!

We have just seen that the power term $1,5^{12}$ gives the value 129.75 instead of 128. How large should the fifth step be, if not 1.5?

If we call this unknown, slightly lower fifth step "x", then we have the equation

$$x^{12} = 128$$

that is $x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x = 128$

The solution to this equation gives the slightly lowered fifth step to be

$$x \approx 1.498.$$

This fifth is called the equal-tempered fifth.

One fifth up from a^1 is e^2. If this is a pure fifth, the frequency of e^2 will be $1.5 \cdot 440 = 660$ Hz. If it is tempered, the frequency is $1.498 \cdot 440$ Hz or 659.12 Hz.

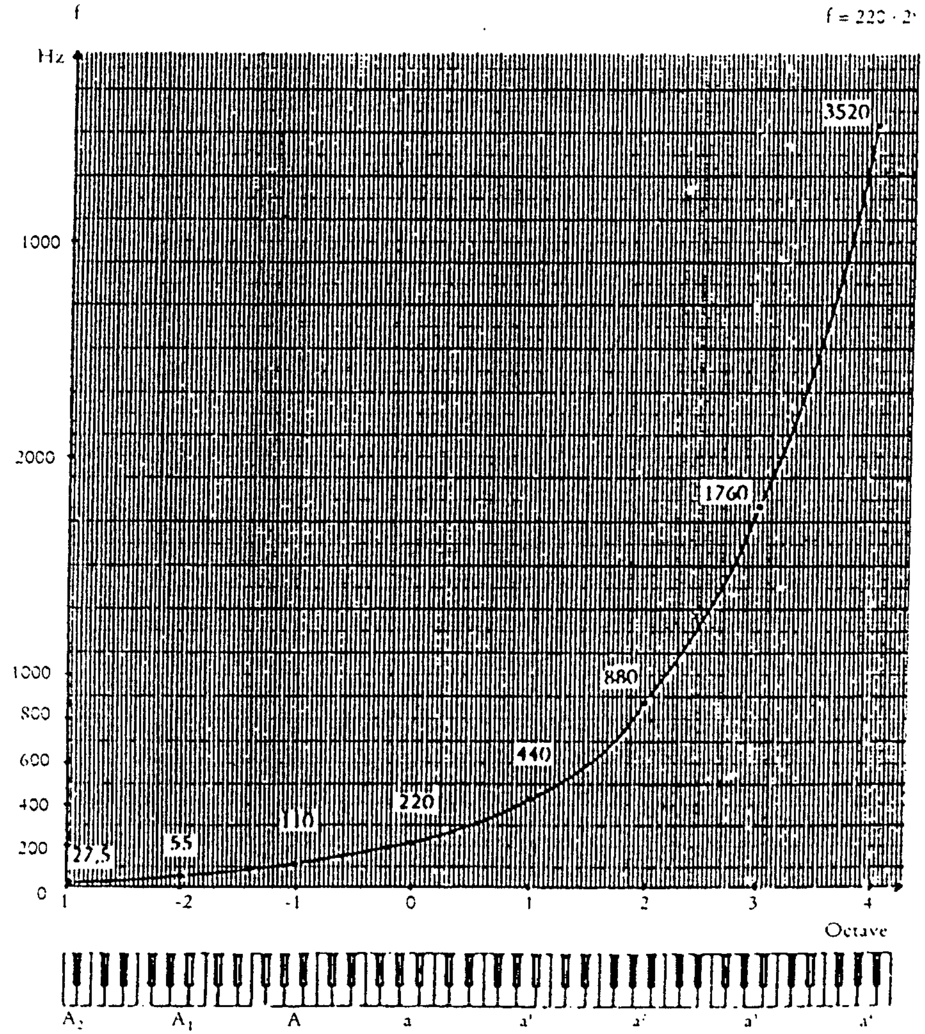


Figure 3.4.10

However, it is not only the fifths which need to be tempered but other intervals as well. And finally it is the half-tones, all of which must be the same size and adjusted so that 12 half-tone steps give an octave. The piano is then "tuned to equal temperature." The equal-tempered piano was introduced during Bach's time.

Thanks to tempering it became possible to go relatively freely between one key and another. With time a new musical form developed

beside the older forms (classical, romantic and others): twelve-tone music. Before tempering, each key was, so to speak, walled-in, and modulating between keys caused problems. As a tribute to tempering Bach wrote the well-known collection of preludes and fugues which carry the title "Das wohltemperierte Klavier" (The equal-tempered piano). Many of us probably recognize the relatively easy-to-play Prelude in C major with which the collection begins. Diagram 3.4.10 illustrates graphically how frequency "grows" on the equal-tempered piano.

3.4.4 Some Psychological Observations

If we hit the keys for A_2, A_1, A, a, a^1, a^3 and a^4 on the piano, we hear "the same" a-tone "repeated" but at steadily higher octaves. The experience is truly such that we may rightly speak of octave *steps*. We identify the octaves with each other and feel we are taking equal steps upward as far as frequency goes.

If we go to the external, physical counterpart, we have the vibrating string, for example, as reality. It is that which puts the air into motion, and the air works in its turn on the ear drum. This physical "teasing" is transmitted thereafter to the inner ear (the hammer, the anvil, the stirrup, etc.) and contributes in a puzzling way to the perception of the tone which was sounded through space.

Let us once again look at the frequency of the vibrating string as we take the octave steps on the A-keys. How does the frequency increase step by step?

	Frequency (Hz)	Frequency Increase (Hz)
A_2	27.5	
A_1	55	27.5
A	110	55
a	220	110
a^1	440	220
a^2	880	440
a^3	1760	880
a^4	3520	1760

We see that the increases get bigger and bigger. From A_2 to A_1 the increase is 27.5, from a^3 to a^4 it is 1760 Hz. How does it come to be that we identify these increases with one another?

In the table above, the increases are calculated as absolute, additive increments. If we instead ask about the *ratios* of the frequencies, the answer is that the ratio for the octave step is always determined by the proportion 2:1. (We could, of course, also say that the *relative* increase is always 100%.)

Our musical ear thus relates to and senses the ratios of the frequencies. Our perception does not follow passively the frequencies which the physical stimulation presents to the musical ear.

How do we react to stimulation levels in other kinds of perception? This is a question which fascinated E.H. Weber (1795-1878), professor of anatomy and physiology in Leipzig.

Weber carried out measurements in what we today call experimental psychology. We will consider a few simple examples here in order to illustrate what Weber discovered:

1. We experience not only tone highness and lowness in music but also tone loudness. We have a special knob or button on the radio receiver for "volume," with which we make the sound louder or softer. More and more has noise become a subject of research. We hear now and then of noise damage to the ear, read of sound levels given in decibels, etc. The question now is: how do we *experience* physical sound level increases? Suppose that we hear a jack-hammer from street work in the distance. After a while there are two jack-hammers in operation simultaneously. If one more jack-hammer, a third, is now put into operation we will experience the sound level increase as less than the previous increase, despite the fact that the increase in sound energy must, of course, be the same as the earlier increase.

But if *two* new drills had started up, so that the number of machines had doubled again, then the new increase would be experienced as equal to the earlier, first increase.

In the case of sound level, too, it is the relative increase in the external stimulation (within certain limits) which is decisive for our comparisons.

2. During the time of classical Greece the stars were divided into so-called orders of magnitude. The name is a misnomer, since it is not a

question of the size of the stars but of their apparent brightness, i.e. of the light intensity which we experience.

The brightest were classified to have first order of magnitude, brighter than those of the second order of magnitude, etc. The Greeks used six orders of magnitude for the stars they could see. Stars of the sixth order are quite dim.

In cameras usually a light intensity meter is built in. Earlier it was common to use a separated light meter held in the hand. When astronomers began to measure the intensity of the starlight reaching earth, they soon found that the step from one order of magnitude to the next lower order of magnitude (i.e. to the next *brighter* group) quite accurately corresponded to a factor 2.5-fold increase in brightness of the light.

In other words, if the sixth order of magnitude has light intensity L, then the intensity of

order 5 is	$2.5L$	=	$2.5L$
order 4	$2.5 \cdot 2.5L$	=	2.5^2L
order 3	$2.5 \cdot 2.5^2 \cdot$	=	2.5^3L
order 2	$2.5 \cdot 2.5^3 \cdot$	=	2.5^4L
order 1	$2.5 \cdot 2.5^4 \cdot$	=	2.5^5L

We once again meet a physical multiplicative factor.

In more recent times orders of magnitude have been defined relative to each other by setting 100 as the ratio of first order light intensity to sixth order intensity. Instead of the five steps from the 6th order up to 1st order, having a factor of

$$2.5^5 \text{ (see table above)} \approx 97.66,$$

they are now given the brightness ration of exactly 100. In order to agree the factor 2.5 changes slightly, becoming about 2.51. (The exact value is the solution to the equation

$$x \cdot x \cdot x \cdot x \cdot x = 100$$

and is written $\sqrt[5]{100}$.

3.4.5 Weber-Fechner's Law or the Psycho-Physical Law

Is it possible to measure inner soul experience, in particular, our perceptions? This and similar questions have undoubtedly been asked and probed into by more than a few minds through the ages.

Galileo strongly advocated the view that science must be based on measurement. His expression, "That which is not measurable must be made measurable" may seem paradoxical if one does not understand it in the sense of: "that which today is not yet measurable must be made measurable in order to be amenable to research." Galileo was one of those who first divided the senses into primary sensory qualities (perception of number, length, shape, etc.) and secondary sensory qualities (experience of color, taste, smell, etc.).

Is it possible to give our experience of tone level or light intensity a certain quantitative or perceptive value? For example, can one set up a perception scale for loudness, as has been done for temperature, length, weight, etc.? Where, in that case, should the zero-point be set? Or must we limit ourselves to comparing different degrees of perception with one another – if that is at all possible?

Such questions engaged Gustav T. Fechner about the year 1850. Fechner had had a dazzling academic career, becoming Professor and Chairman of the Physics Department at the famous University of Leipzig at the age of 33 (in 1834). But he apparently overworked himself in a variety of ways and left his teaching chair in 1839.

Fechner attacked his new research with great energy. After publishing a few short papers in 1858 and 1859, he gave out his book "Elemente der Psychophysik" in 1860, a famous work on "the exact science of the functional relation between body and soul."

Did Fechner succeed in capturing perceptions of the soul as quantities, on scales? The answer is surprisingly more-or-less "yes," if we limit ourselves to a few areas, primarily tone level, loudness and brightness – those areas we have already touched upon. We have actually already learned that which Fechner stated in his psycho-physical "law", that perceptive level grows with equal steps when the physical stimulus grows multiplicatively in equal steps; or, expressed differently, when the stimulus grows relatively in equal steps.

Such a "law" cannot, of course, be expected to apply over an unlimited range of perception. For one thing, the stimulus (tone frequency, loudness or light level) must be strong enough that we clearly perceive something above the so-called stimulus threshold. Secondly, the stimulus must not be so strong that we begin to approach the level of pain or other reactions which disturb our senses.

We cannot here go into a thorough study of Weber-Fechner's law and the criticism of it. Instead we will use a graphical presentation to give a clearer picture of that which Fechner asserted.

In Figure 3.4.10 earlier, where a curve is drawn of tone frequency, we used common linear scales on both axes. How does the diagram look if we replace the vertical frequency scale with a *logarithmic* scale but let the horizontal x-axis remain linear? (For reference to the concept of logarithmic scale, see Section 3.4.2.)

For this we mark off the frequency values for the octaves with *equal* spacing along the vertical axis; the length of the interval can be chosen arbitrarily. How do the points now lie?

We set out the points for -3 and 27.5, for -2 and 55, etc. and finally for 4 and 3520. We note that these points now lie in a straight line. For every step sideways the point moves vertically by the same constant amount, equal to the distance chosen for the vertical

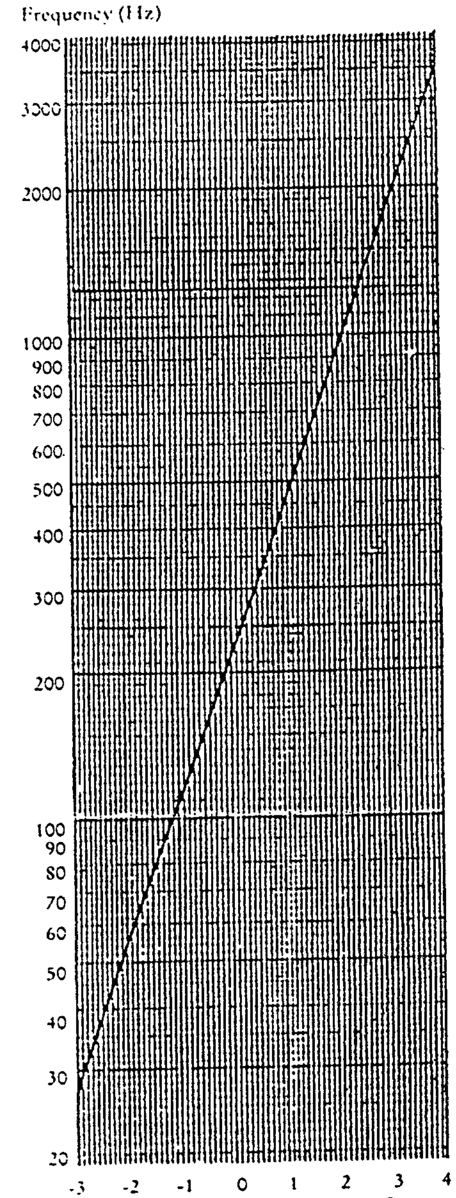


Figure 3.4.11
Graph on logarithmic paper

axis octave markings (see Figure 3.4.11). The octaves correspond to the equal distances between points on the line.

Noise research has grown in importance in recent years, especially that aimed at making homes and places of work more humane, so that people at least avoid damage to their hearing. Measurement of loudness was adapted to the fact that our perception of sound level basically follows the Weber-Fechner law: the loudness was set according to our perceptions and made logarithmic. The scale's unit is called decibel (db). In order to understand how the decibel scale is constructed we shall here make up a new scale for frequencies. We use a^1 as a base and call its frequency 50 "Euler" (honoring the mathematician Leonard Euler, who presented the psycho-physical material concerning tone intervals and frequency). Further, we let 10 Euler correspond to one octave step.

Then a^2 will have the frequency 60 Euler, a^3's frequency will be 70 Euler, etc. Going in the other direction we get

$$a = 40\,E \quad A = 30\,E \quad A_1 = 20\,E \quad A_2 = 10\,E$$

and arrive at the following table of comparisons:

Tone	Frequency in Hz	Frequency in E
a^2	880	60
a^1	440	50
a	220	40
A	110	30
etc.		

What frequency in Hz would correspond to the Euler frequency 0? Apparently that frequency which is one octave under A_2, so we get

$$0\,E = 0.5 \cdot 27.5\text{ Hz} = 13.75\text{ Hz},$$

and further

$$-10\,E = 0.5 \cdot 13.75\text{ Hz} = 6.875\text{ Hz, etc.}$$

The Euler scale number goes toward negative infinity, as the number of vibrations per second decreases toward zero (see Figure 3.4.12). We have by then left the world of tone, since our perception of tone stops below about 20 Hz. The decibel scale for loudness is constructed in a manner completely analogous to our construction of the Euler scale here. 10 decibels means a doubling of loudness.

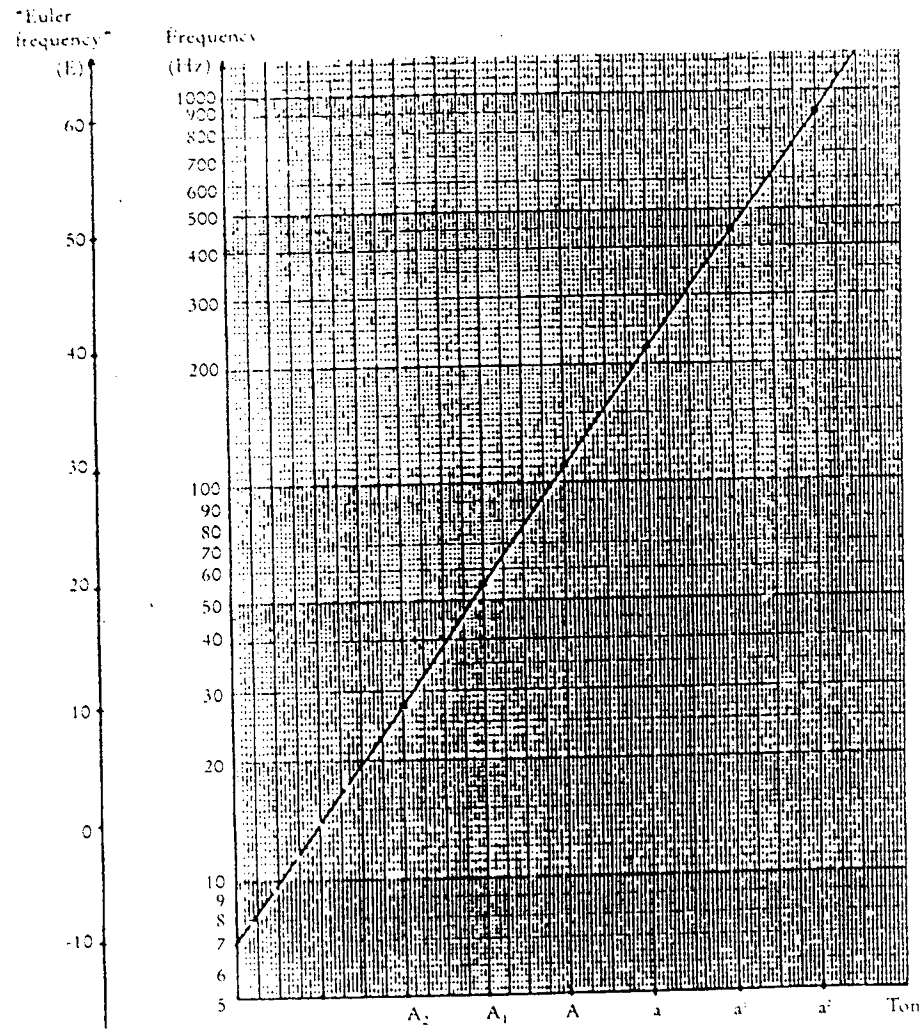


Figure 3.4.12

In summary of this look into our soul's activity during the perception of tones, sound level, and light intensity, we may say, graphically speaking, that we unconsciously transform a linear scale to a logarithmic one. Those somewhat familiar with the concept of logarithms will understand another formulation of the Weber-Fechner law: in the process of perception we take the logarithm of the intensity of the stimulus.

3.4.6 *A Little Population Mathematics*

When youth are called into the army, their intelligence, among other things, is tested. Some of the test problems have been arithmetic of the following type:

Example 1 A numeric sequence begins with the numbers
3, 6, 9, 12 ...
What comes after 12?

Example 2 What number comes after 3, 6, 12, 24, ... ?

I have chosen two simple examples. We answer with 15 and 48 respectively. More correct would be to say: the *simplest* continuation gives 15 and 48. In the first example we observe a constant increase (3), in the second, a doubling of the numbers, step by step – the sequence has the multiplicative factor of 2. We might remind ourselves here of Moser's circle subdivision problem in Section 3.1 where we successively obtained the numbers 1, 2, 4, 8, 16 and were inclined, perhaps even took it as obvious, to predict 32 as the next following number. In actual fact the next number after 16 was 31.

If in advance we state that we consider only sequences with constant increases (or decreases), or with constant multiplication factors, we may unambiguously state the continuations in the following examples:

Example	3a)	7, 11, 15, ...
	3b)	23, 16, 9 ...
	3c)	80, 20, 5 ...
	3d)	2, -6, -14, ...

(It would have been sufficient to give only the first two numbers.)

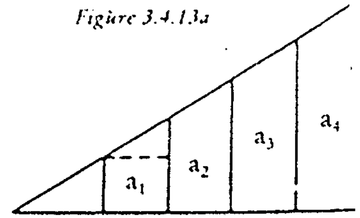
The numbers are 19, 2, 1.25, and -22 respectively.

Sequences with constant multiplication factor (Examples 2 and 3c) are called geometric sequences. If the sequence on the other hand has a constant increase (Examples 1 and 3a) or a constant decrease (i.e. a negative increase, as in Examples 3b and 3d) then it is called an *arithmetic* sequence.

Figures 3.4.13a and b illustrate these two types of sequences geometrically (see also Section 3.7.2).

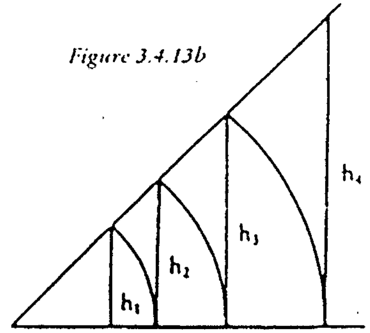
How does the population of a city, a country, a continent, or of the world grow?

Figure 3.4.13a



Arithmetic sequence

Figure 3.4.13b



Geometric sequence

This is always a current question. More than anyone else Thomas Robert Malthus (1766-1834) has become known for his investigation of this question.

Malthus, who studied at Cambridge, was ordained as a priest in 1798 and that same year gave out a book called "An Essay on the Principle of Population As It Affects the Future Improvement of Society," which brought sharp criticism and started a lively public debate in England.

The book was published anonymously and directed toward an optimism for the future which had come to expression through the anarchist Godwin. The Liberals had introduced poor-laws in England in 1796 to create a better society. Such a measure would, according to Malthus, son lead to worsened conditions for the poor. The supposed improvements would stimulate increased nativity and after a time the growth in the numbers of poor would lead to new and even more difficult times of need.

Malthus noted a number of factors which, according to him, acted to limit the population increase: poverty, sickness, and war. If such inhibitors do not exist the population continues to increase geometrically, i.e. the number of people forms a geometric sequence, when equal points in time are taken.

Malthus' starting point was that "a population of a thousand million doubles just as easily in 25 years as a population of a thousand."

> In the Northern States of America..., the population has been found to double itself, for above a century and a half successively, in less than twenty-five years. — In the back settlements, where the sole employment is agriculture, and vicious customs and unwholesome occupation are little known, the population has been found to double itself in fifteen years.
>
> But the food, which shall support the increase of the greater population, is by no means obtained easily.

Here we come to the main thought in Malthus' presentation: while population increases geometrically, food production can only increase arithmetically.

> ... supposing the present population equal to a thousand millions, the human species would increase as the numbers 1, 2, 4, 8, 16, 32, 64, 128, 256, and subsistence as 1, 2, 3, 4, 5, 6, 7, 8, 9. In two centuries the population would be to the means of subsistence as 256 to 9; in three centuries as 4096 to 13; and in two thousand years the difference would be almost incalculable.[2]

In other words: if a population of 1 billion people today divides up 1 unit of food stuff among themselves, then 256 billion people 200 years later would have 9 units of food to live upon. Poverty would increase enormously – if none of the inhibiting factors of hunger, sickness, and war were to slow the increase down.

JAPAN

Year	Mill. inhab.
1850	27
1870	32
1890	40
1900	44
1910	50
1920	56
1930	64
1940	73
1950	84
1960	94
1970	104
1980	117

Criticism led Malthus to collect empirical evidence for a new edition of the book. It was published in 1803 after Malthus had undertaken study trips to Germany, Sweden, Norway, Finland, and Russia. He stood firm by his basic hypothesis in the new edition but pointed out that other factors than hunger, sickness, and war could hold back the population increase and improve the supply of the means of subsistence. As examples of favorable positive factors to raise the minimum level of existence Malthus suggested late marriage, voluntary restraint, and the formation of new habits.

Does the hypothesis that a population grows geometrically in equal periods (as long as living conditions are good) hold true for any country or region of the world? Let us look at Japan:

[2] The meaning here must be that the fraction $2^n/(n-1)$ would be incredibly large.

Has population increased in geometric sequence in Japan over the span 1850-1970 or during any part of that period? (See table). Diagram 3.4.14 shows graphically how Japan's population has increased. In order to study the growth of population we will acquaint ourselves with a graphical test method which reveals very simply whether a growth is geometric or not.

Semi-Log Diagram

In Section 3.4.2 we introduced the concept of a logarithmic scale. On such a scale a geometric series is compressed so that the numbers in the geometric series are put down with equal spacing. What does the curve of a geometric sequence look like on a diagram where the horizontal scale for the order in the sequence is linear, while the vertical scale for the numbers themselves is logarithmic? Such a diagram is said to be semi-logarithmic. Graph paper is available with pre-printed scales to use. This paper is called semi-log graph paper.

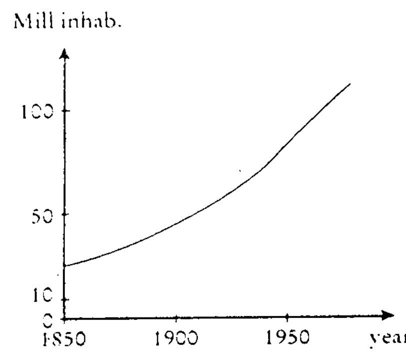


Figure 3.4.14

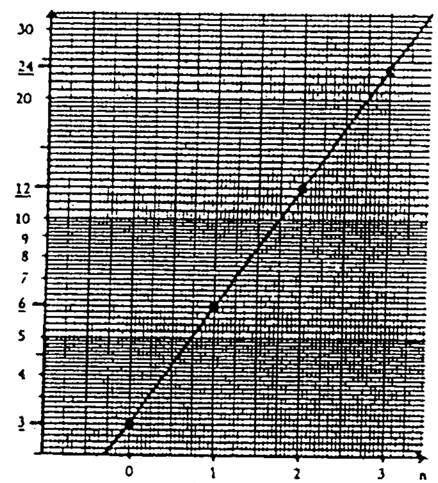


Figure 3.4.15

The result may be seen in Figure 3.4.15. The curve is, quite simply, a straight line. Geometric growth gives a straight line in a semi-log diagram. This is a practical method of graphically testing whether or not a given growth is geometric.

In Figure 3.4.16 we see the population of Japan on a semi-log plot. The diagram in large measure lends support to the hypothesis that

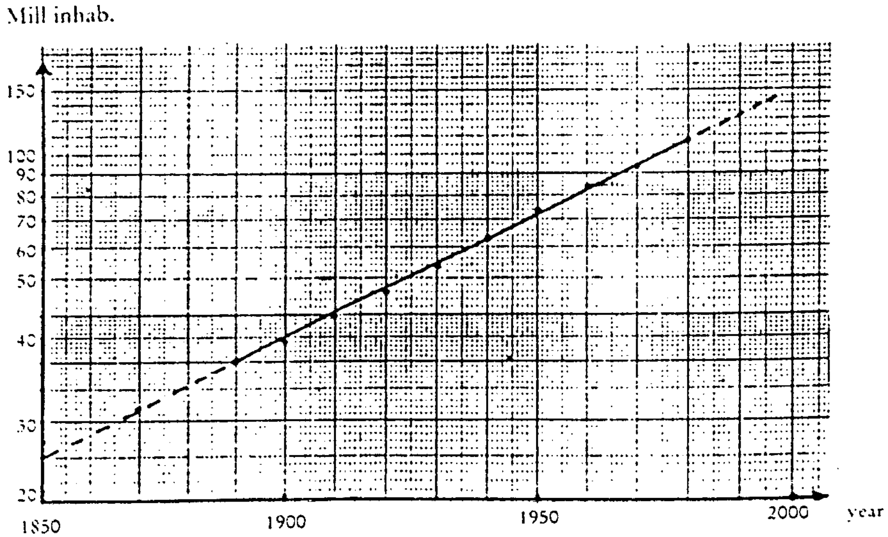


Figure 3.4.16

growth has been geometric. The results are not equally convincing if one studies, in the same way, the population of the United States for the same period...

Malthus is still very timely. In a newspaper account from the International World Population Conference in Bucharest 1974, could be read, among other things, the following:

> The ghost of the English doomsday prophet Malthus hung over the congregation after the opening speech.... Malthus became a weapon at the conference for those who wanted to get at the developmental pessimists and defenders of the privileged. China and every one of the attending African and Latin American nations quickly shot at what they thought were "neo-Malthusian" tones in the plan. This group considered that the plan, developed by the United Nations Commission on Population, placed far too much emphasis on family planning.

It goes without saying that Malthus can spark strong interest in a tenth-grade class.

3.4.7 Exercises

1. What are the fourth and fifth terms in the arithmetic sequence

   1   3,4   5,8 ... ?

2. What are the fourth and fifth numbers in the geometric sequence

   5   8   12.8 ... ?

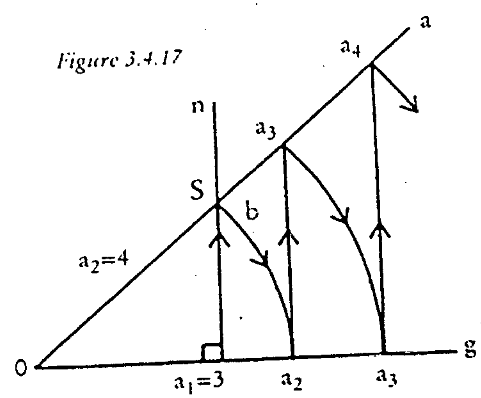
3. The numbers 12, 6, and 3 are the beginning of a decreasing geometric series. Which numbers follow?

4. The angles of successive swings of a pendulum form a decreasing geometric sequence. Determine the angle of the sixth swing if the first three swings are 15°, 12°, and 9,6°.

5. Figure 3.4.17 shows a graphical method of constructing a geometric series using only compass and straight edge. The n-lines are drawn vertically from the base line g up to the line a. Circular arcs are rotated from a down to g, using O as pivot center. The distances from O to the points along g now form a geometric series. The first two numbers in the sequence in the figure are 3 and 4. First the distance 3 cm is marked off from O to the line n with the aid of a compass. This intersection along n determines a's slope. Line a can now be drawn, and the method of constructing a geometric series with 3 and 4 as starting numbers may proceed.

   Letting 5 and 8 be starting values, determine graphically the fifth number in the sequence. Compare with the results in Exercise 2 above.

Figure 3.4.17



6. A mirror is found to reflect 95% of the incoming light. What percent of the light is lost if light is reflected successively in 4 such mirrors?

7. Russia's population grew as follows over the period 1850-1970:

Year	1850	-70	-90	1900	-10	-20	-30	-40	-50	-60	-70
Million people	61	75	99	111	140	134	156	174	181	214	245

Illustrate this growth with a plot. What can be observed between 1910 and 1920?

Now illustrate the population growth on a semi-log plot. If the curve obtained is a straight line for any period of time, it indicates that population growth was exponential (geometric) during that time period.

3.5 The Step from Arithmetic to Algebra

3.5.1 Why "Algebra"?

We use the word "algebra" here in the limited sense of arithmetic with letters and "arithmetic" in the sense of numerical calculations.

Many pupils find calculation with letters a, b, c, x, y, etc. to be abstract. For the majority numerical arithmetic is considerably more concrete, and it is truly an important pedagogic problem of how to introduce arithmetic with letters. Often — for many pupils — algebra seems unnecessary. The teacher will certainly hear the question, "What's this good for? Does it have any use?"

Such a reaction is quite natural if the pupils have not at a rather early stage — as soon as the prerequisites exist — had an experience of the "power of algebra."

Let us proceed directly to a problem concerning numbers:

We choose four arbitrary numbers — for simplicity's sake one- or two-digit numbers — and write them beside one another with a little space between, on the blackboard. Suppose we have chosen the numbers 3, 11, 4, and 17. We now add adjacent numbers two at a time and write their sums one row down in the spaces in between:

```
3      11      4      17
   14      15     21
```


We repeat this procedure and get

```
3      11      4      17
   14      15     21
       29      36
```

Finally we do the last sum, which we call the *bottom number*:

```
3      11      4      17
   14      15     21
       29      36
           65
```

Our choice of initial numbers led to an *odd* bottom number.

The question now is: can we find some simple rule with the help of which we can predict whether the bottom number will be odd or even, as soon as we know the four starting numbers?

For example, if they are 1, 9, 16, and 8, will the bottom number be odd or even?

Without exception the students add up the sums to see what bottom number appears. They choose their own examples to try out, and I ask them to be alert concerning the results so that we can come up with a theory, at least a guess, from our collection of examples.

Someone soon points out – perhaps immediately – that if all the given numbers are even, then the bottom number will also be even. The same applies if all the numbers are odd, someone adds. And now groups of pupils go exploring in different directions.

Some groups investigate the effect of the number of even numbers in the beginning set of four. Other groups investigate if the sum of the given numbers is of any use. Someone says that they have discovered that the bottom number is even, if the sum of the initial numbers is even. Will the bottom number be odd, in this case, if the initial numbers' sum is odd? The search becomes more intensive. The conviction that the sum of the initial numbers determines the even-oddness of the bottom number grows stronger. "All our examples thus far agree with that," point out some pupils and they begin to consider the result as certain.

We gather together again as a class and listen to this theory about the sum of the initial numbers. "How many examples would you deem it necessary to exhibit, in order for the theory to be proved?" I once asked in a ninth-grade class which had not seen the problem before and which was not especially advanced in algebra. The class took a long "think" and first after a pause came several cautious opinions. Those who answered believed that four or five examples would be sufficient. This became a lovely prelude to a discussion of what is meant by a proof and of the value of a few examples versus a great number of examples. It soon became clear to the pupils that not even a million examples which all agree are sufficient as a proof – despite the strong conviction we might then have. We even got into a discussion of the value of experiment in science versus "experiment" in mathematics.

3.5.2 *We "Experiment" Further*

There also came up examples of experiences with bottom numbers for the case when some of the given numbers were even (or odd). Before we seriously took on the task of proving one or the other of our theories we were tempted to "experiment" using 5 starting numbers. If we had 5 given numbers would examples also indicate that the sum of the numbers determines the bottom number's quality? Several examples at first seemed to confirm this, for example 1, 5, 8, 9, 11 with the row sum 34:

```
1     5     8     9     11
   6    13    17    20
     19    30    37
        49    67
          116
```

If the bottom number is determined solely by the top row sum, then the "even" quality should be maintained if we exchange places in the top row, e.g. the 1 and the 8:

```
8     5     1     9     11
  13     6    10    20
     19    16    30
        35    46
           81
```

But here the bottom number is odd!

We have found a counter-example to the theory that the bottom number's even-oddness is determined by the sum of the five given numbers. One single example is enough to overthrow a theory!

But what might the rule then be, if there is a simple rule?

There were still hopeful students who sought for some special rule. A few came with the logical and promising idea that we could try to relate the case with 5 numbers back to the earlier case with four numbers, since the first row of additions gives 4 new numbers. But then studying 6, 7, 8, or more given numbers with successive "backtracking" downwards would be laborious to carry out and difficult to summarize.

Why not try putting letters in place of the numbers?

We decided to go back to the case with 4 given numbers and to call these a, b, c, and d. We formed the sums row by row and obtained the following triangle:

```
a          b          c          d
     a+b        b+c        c+d
          a+2b+c     b+2c+d
              a+3b+3c+d
```

We have thus obtained the bottom sum: $B = a + 3b + 3c + d$.

Does the even-odd quality of the sum $s \doteq a+b+c+d$ determine the quality of this expression, B? We compare B and s and find that

$$B = s + 2b + 2c$$

$$\text{or} \quad B - s = 2(b + c) \qquad (1)$$

Here we see that $2(b + c)$ is always an even number — it is divisible by 2. Equality (1) tells us that B and s are *simultaneously* even or odd, i.e. either both are even or both are odd. We have hereby come to a clear and proven conclusion for the case with 4 initial numbers:

> the bottom sum is even or odd when the sum of the initial numbers is even or odd, respectively.

The class wants now to try out this effective method for the case when we have 5 numbers to start with. What bottom number will we then get, and what conclusion will we be able to draw?

We must now start with 5 letters, each representing a number, any letters we like, say a, b, c, d, and e. Without great difficulty we come up with the bottom sum

$$B = a + 4b + 6c + 4d + e$$

Now what kind of rule can we get out of this expression? It turns out to be a difficult question for the majority. At this point many simply copy, without a closer thought, the comparison between B and s which we did earlier and which led to equality (1). They arrive at the relation

$$B = s + 3b + 5c + 3d$$

but come no further, since the sum $3b + 5c + 3d$ is even for even numbers and odd for odd b, c, d.

But a glance shows us that the rule is hidden within B itself:

$$B = a + 4b + 6c + 4d + e$$

which says that $B = a + e +$ an even number, namely $2(2b + 3c + 2d)$.

With this it is clear that $a + e$, i.e. the sum of the outer numbers determines the bottom number's quality when we have five numbers initially.

The rule is simpler when we have 5 numbers than when we have 4!

The class usually gets so involved that they want to go on: does the rule get even simpler when we have 6 numbers to start with? Or do we get a rule similar to the case with 4 numbers?

The results are once again surprising! And give further opportunity for wrestling with questions towards a more general investigation. Several examples of different students' investigations and partial results are found in the exercises following.

But *one* observation shall be noted here, before we leave the bottom numbers:

2 starting numbers give	$B = a + b$
3 " " "	$B = a + 2b + c$
4 " " "	$B = a + 3b + 3c + d$
5 " " "	$B = a + 4b + 6c + 4d + e$
6 " " "	$B = a + 5b + 10c + 10d + 5e + f$
etc.	

Do we perhaps recognize the numbers which appear in the respective rows?

$$\begin{array}{ccccccccccc} & & & & 1 & & 1 & & & & \\ & & & 1 & & 2 & & 1 & & & \\ & & 1 & & 3 & & 3 & & 1 & & \\ & 1 & & 4 & & 6 & & 4 & & 1 & \\ 1 & & 5 & & 10 & & 10 & & 5 & & 1 \end{array}$$

Why, yes, they are actually the same as the numbers in Pascal's triangle! Can we obtain B so simply, by going to Pascal's triangle? (Exercise 2).

3.5.3 *The Rule of 9's*

Pupils usually learn a rule in grade 4 or 5, of which many of us have heard. It is the answer to the question: when is a whole number evenly divisible by 9? The rule says: a number is divisible by 9 when the sum of the digits is divisible by 9 – and only then. (The corresponding rule applies for 3).

A couple of examples: Is 2169 divisible by 9?
Sum of the digits: $2 + 1 + 6 + 9 = 18$, divisible by 9.
the number 2169 is therefore divisible by 9.

And if we take 31,478?

Digit sum: $3 + 1 + 4 + 7 + 8 = 23$, *not* divisible by 9, so neither is the number.

How can we prove such a rule?

Let us first limit ourselves to 3-digit numbers. We suppose that the 3 digits in the number are a, b, and c. How would the number be written? abc? No, since abc in algebra means $a \cdot b \cdot c$ – we don't write out the multiplication sign between letters. 538 certainly does not mean $5 \cdot 3 \cdot 8$ but rather 5 *hundred thirty* (three ten) eight, i.e.

$$500 + 30 + 8 \text{ or more clearly } 5 \cdot 100 + 3 \cdot 10 + 8$$

How shall we then write our number?

Correct: $n = a \cdot 100 + b \cdot 10 + c$
or more nicely $n = 100a + 10b + c$

And the sum of the digits? Correct answer: $s = a + b + c$

How can we compare n with s?

We can perhaps write them underneath each other:

$$n = 100a + 10b + c$$
$$s = a + b + c$$

and see that the *difference* between them is $n - s = 99a + 9b$

$$\text{or} \quad n - s = 9\,(11a + b), \qquad (2)$$

a number which is *always* divisible by 9.

We "solve" for n and get from (2)

$$n = s + 9\,(11a + b) \qquad (3)$$

What does equality (3) tell us?

It says: if s is divisible by 9, then n will also be divisible by 9. But is n divisible by 9 *only* when s is divisible by 9?

Formulated differently: when n is divisible by 9, must s also be so?

We return to (3) and solve for s, getting

$$s = n - 9\,(11a + b) \qquad (4)$$

Equality (4) tells us that s is divisible whenever n is – and only then. With this we have completely proven the rule for 3-digit numbers. Since the argument may be carried through analogously with an arbitrary number of digits we understand that any number is divisible by 9 whenever its digit sum is divisible by 9.

3.5.4 The Example as Teacher

One day on the Uppsala-Stockholm commuter train I was witness to a dispute between two young men, university students — let us call them A and B — who had differing opinions as to the speed of the train. "Why don't you just calculate the velocity; you've got a calculator?" said A. "The distance is 66 km, and the travel time is 40 minutes." "You'll see I'm right," said B and took out his calculator. He began to punch keys. But no result seemed to be forthcoming. We passed Marsta, then Upplands Vasby and came into the Solna tunnel near Stockholm. I began to guess that B had searched his memory in vain looking for a formula containing distance, time, and velocity (s, t, and v) and that now he was simply experimenting with his calculator.

In such a situation one may come to one's own aid with the help of a simple example, where the three qualities can be seen in relation to each other.

For example: how far does a train travel in 3 hours if it goes 80 km/hour? The train obviously covers a distance of $80 \cdot 3 = 240$ km.

From this simple example we see directly the relation

$$\text{distance} = \text{velocity times time}$$

(assuming that velocity is constant, of course).

In a short version: $\quad s = v \cdot t \quad$ or $\quad s = vt \qquad (1)$

We need not store this formula in memory. We can have faith in our imagination to find a concrete example and in our thinking ability to extract the formula from the example, once again.

By the very act of writing formula (1) we take the step from numeric example to algebra, limited however, in the sense that we use letters as abbreviations for quantities which occur in the formula. What do we do when velocity is unknown as in the dialogue above?

One might then take the example that a train covers 90 km in 2 hours. How fast did it go?

Naturally 90/2 = 45 km / hour (most likely a freight train).

It matters not at all if the results of our home-made examples are a little far-fetched. The main thing is that we set our thinking in motion through the choice of concrete, numerical values.

This time we find that

$$\frac{\text{distance}}{\text{time}} \quad \text{or} \quad v = \frac{s}{t} \qquad (2)$$

If one brings to mind that speeds for trains and cars are usually given in km / hour, one sees that it must be a question of dividing a distance by a time. Those who are used to working with letters need not formulate more than the first example: from (1) we can easily derive formula (2). If in fact we divide both sides of (1) by t we will get

$$\frac{s}{t} = \frac{v \cdot t}{t} \quad \text{and after reducing} \quad \frac{s}{t} = v.$$

If we instead divide (1) by v on both sides, we get a third relation,

$$\frac{s}{v} = t.$$

Formulas (2) and (3) are just reformulations of the basic relation in (1).

There are many examples of relations of this kind. A few are:

Area of a rectangle = length x width	$A = l \cdot w$
Mass = density x volume	$m = d \cdot V$
Voltage = resistance x current	$E = RI$ (Ohm's law)
Cylinder volume = base area x height	$V = B \cdot h$

An example can be of value not only when we are looking for a formula. It can sometimes also give us the solution to a problem which is broader than the original task.

Let us look at the following problem, which we imagine using in a sixth or seventh grade class:

What kind of sums do we get when we add three consecutive integers? For example, 23 + 24 + 25, 6 + 7 + 8, etc. We ask each pupil to give at least one example, and take the time to write down all the totals which the pupils obtain. It may happen now and then that an incorrect sum gets in, but this only makes the exercise more interesting. The list might begin like this:

45, 95, 12, 156, 51, 15, 78, 222, ...

Do these numbers have anything in common?

The class soon discovers that the sums are divisible by 3 and are entirely convinced that they *always* can be divided by 3. Might there not be exceptions? How could we prove the divisibility?

At this point it is usually a great help to examine *one* example thoroughly. Someone has taken 13 + 14 + 15 and obtained 42. Might we from the very arrangement itself, 13 + 14 + 15, be able to predict that the total is divisible by 3?

The class begins thinking. We wait with our answers so that everyone has time to think. Soon several students quietly come up and show on paper an idea that they have. For example, 15 lies just as much above 14 as 13 lies under. The sum is therefore $3 \cdot 14$, that is, it is divisible by 3. The moment we see the correctness of that idea we understand that it applies *generally:* the same argument can be applied to any sum of 3 consecutive integers. The sum must therefore always be divisible by 3.

We see here that thinking about a particular example can be so fruitful and enlightening that we arrive at a general result.

It is no great task afterwards to clothe the reasoning in algebraic dress: letting the middle number be N, the sum is

$$N - 1 + N + N + 1 = 3N$$

The class might now be given the task of investigating if similar results can be obtained for sums of four, five, or six integers.

3.5.5 *Magic Hexagons*

In the lower grades the pupils have eagerly added rows, columns, and diagonals in "magic squares" and perhaps themselves constructed a simple magic square, for example a square with 9 boxes containing the numbers 1 through 9, in which the rows and the columns and even the diagonals all add up to the same sum.

8	1	6
3	5	7
4	9	2

Albrecht Dürer constructed a square with 4 x 4 boxes, containing the numbers 1 through 16, which, by the way, was done in such a way that two adjacent boxes in the bottom row gave the year of construction, 1514.

16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

There are many books available which can acquaint us with magic squares of even larger size. Can one construct a square with only 2 x 2 boxes containing four whole numbers? That the numbers 1, 2, 3, and 4 are not a solution requires no great effort to discover. But can we succeed with four other integers? We return to this problem in Exercise 3.

An American office worker, Clifford W. Adams, posed himself the problem of constructing magic hexagons (6-sided figures). We call a single hexagon a hexagon of the first order, a group of 7 hexagons as in Figure 3.5.1 is called a hexagon of the second order and a group of 19 hexagons as in Figure 3.5.2 a third order hexagon. Is it possible in Figure 3.5.1 to put the first seven numbers 1, 2, 3, ... 7, and in Figure 3.5.2 to put the 19 numbers 1 to 19, so that one and the same sum is obtained when one adds the numbers in the boxes horizontally, and along each of the two directions for diagonals? (Figure 3.5.3)

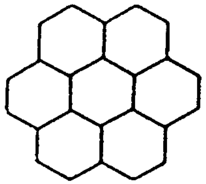


Figure 3.5.1
Hexagon of the second order

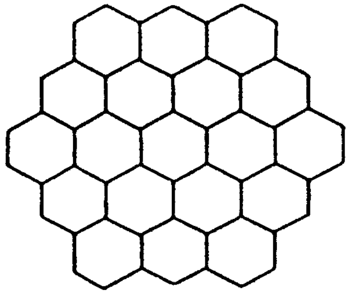


Figure 3.5.2
Hexagon of the third order

If we let a,b, and c be three numbers placed in adjacent boxes in a second order hexagon, as in Figure 3.5.4, then we are to have the row sum a + b be equal to the diagonal sum b + c. Is this possible? What conclusion can we draw from the equality

$$a + b = b + c?$$

The equality applies only if a = c, which means that two boxes contain the same number. Since no number may appear more than once, we see that a magic hexagon of the second order does not exist. But perhaps one exists of third order, with 19 boxes?

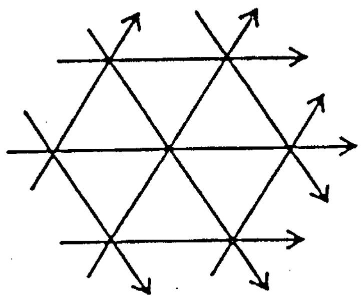


Figure 3.5.3

Adams began his investigations, according to an article in *Scientific American* (No. 8, 1963), the year 1910. He proceeded by trying to place the numbers 1, 2, 3, ... 19 in the boxes in different ways. For a long time he did not succeed, but in 1957, after 47 years, the retired Adams found a solution. It must have been a shock for him when some time later he discovered that he had misplaced the paper with the

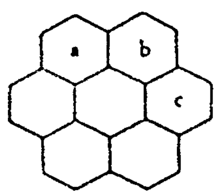


Figure 3.5.4

magic hexagon he had constructed and that he would not remember the placement of the numbers. Should he begin again trying out different combinations, or should he try to find the paper? He looked long and in vain for the paper, but finally found it five years later! The solution then came to Martin Gardner's attention. (Gardner was the editor of the mathematics corner in *Scientific American*.) He turned to W. Trigg with a request for a general mathematical investigation concerning the existence of magic hexagons of arbitrary order. Trigg proved in 1963 that no magic hexagon of higher order than 3 exists. Perhaps it was good intuition which saved Adams from the Sisyphean task of looking for a magic hexagon of order 4.

3.5.6 *Mathematical Induction*

Example 1: When Galileo studied how far a ball rolls down an inclined plane during a specified time, or how far a marble falls in the air during a given time, he came to the conclusion which he formulated in the following way:

The distances which the ball rolls or falls during successive, equally long time intervals are proportional to each other as the odd integers:

$$1 : 3 : 5 : 7 \ldots$$

If a ball rolls 1 "bit" during the first second, then
it will roll 3 "bits" during the second second, and
5 "bits" during the third second, and
7 "bits" during the fourth second
etc. (Fig. 3.5.5)

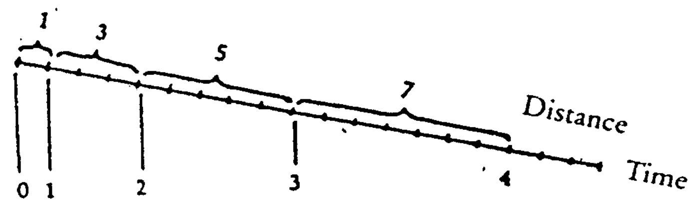


Figure 3.5.5. Galileo's rolling ball

If we now add up the distances from the starting point, we find that the total distance traveled is

1 "bit" in 1 second
$1 + 3 = 4$ "bits" in 2 seconds
$1 + 3 + 5 = 9$ "bits" in 3 seconds
$1 + 3 + 5 + 7 = 16$ "bits" in 4 seconds
etc.

The numbers which determine the total distance are thus 1, 4, 9, 16, ... These numbers are *squares* of 1, 2, 3, 4, etc., i.e. of the numbers which specify the times. Does this pattern continue? And if so, out to infinity? If we leave mechanics and formulate the question purely mathematically, our problem is to find out:

Does the sum of the series

$$1 + 3 + 5 + \ldots + \text{a last odd number}$$

make a perfect square?

$1 + 3$ gives 2^2, $1 + 3 + 5$ gives 3^2, etc. It appears as if the sum is the square of the number of terms.

For example, we would expect

$$1 + 3 + 5 + 7 + 9 + 11 = 6^2 \quad \text{(since there are 6 terms)}$$

We check it: the sum is actually 36.

How can we prove our supposition?

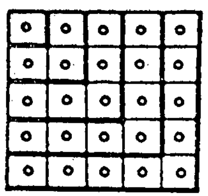


Figure 3.5.6

The Greeks found a beautiful geometric proof: Figure 3.5.6 shows a square which is made up of a number of L-shaped right angles ("gnomons"). The number of points in the right angles corresponds to the successive odd numbers. Their sum is the number of points in the whole square. The figure shows directly that $1 + 3 + 5 + 7 + 9 = 5^2$.

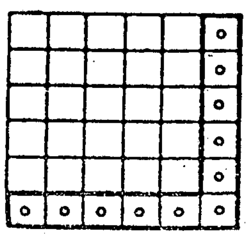


Figure 3.5.7

What happens now if we add 11 more? A new, larger right angle piece gets added on, and we see with the help of Figure 3.5.7 that this piece together with the old square makes a *new square* which corresponds to the sum 6^2. If we now continue to add on larger and larger angle pieces, do we always get the next largest square?

The new right angle has two points more than the previous. These two extra points are just

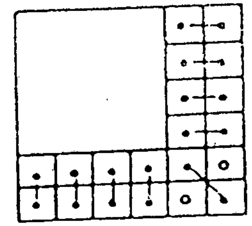


Figure 3.5.8

what are needed so that the larger right angle can go round the corner of the previous square (Figure 3.5.8).

That brings us to the insight that a square plus a right angle makes the next larger square, step by step.

We can now draw the conclusion that the sum of the old integers always gives a perfect square, taking as many integers as we like.

In summary we may say:

(1) $1 + 3$ gives the square of the number 2.

(2) Adding on the next following odd number gives the next larger perfect square.

(3) This can be repeated endlessly.

(4) The sum of the successive odd integers always gives the square of the number of terms.

Concentrating the proof to *one step* which can be repeated an unlimited number of times, starting from a proven beginning point, is the kernel of the elegant means of proof called mathematical induction.

This method of proof was popularized by Blaise Pascal, who even in his early years gave evidence of exceptional mathematical talent (see also 3.2.3). The point in the induction method of proof is that one need not look for a direct proof. In our example we avoid the task of directly summing $1 + 3 + 5 + 7 + \ldots$

If we wish to carry out inductive proof purely arithmetically for our odd number series, we can begin by *supposing* that the sum of n terms

$$1 + 3 + 5 + \ldots + (2n - 1) \quad \text{is } n^2. \tag{1}$$

We know that this is correct for $n = 1$. (2)

If we add now the next following odd number, $2n + 1$, to series (1), the sum will be, according to our supposition

$$n^2 + 2n + 1.$$

Is this a square? Yes, since $n^2 + 2n + 1 = (n + 1)^2$.

If (1) is true, then it is also true that

$$1 + 3 + 5 + \ldots + (2n + 1) = (n + 1)^2 \qquad (3)$$

The *step* of adding the next higher odd integer thus gives once again a sum which is the square of the number of terms. We can now go back to the beginning (2) and carry out the step from (1) to (3) as many times as we like. The induction is complete.

Example 2: Finding a formula for the sums of squares is much more difficult:

$$1^2 + 2^2 + 3^2 + \ldots = ?$$

The first four sums are

$$1^2 = 1$$
$$1^2 + 2^2 = 5$$
$$1^2 + 2^2 + 3^2 = 14$$
$$1^2 + 2^2 + 3^2 + 4^2 = 30$$

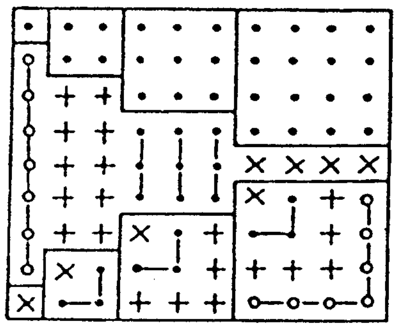


Figure 3.5.9

Compare with the figurated numbers in Figure 3.1.4.

It is interesting that the Babylonians succeeded in finding a formula for the sums of squares and also gave a geometrical proof by induction for the formula. The method is in principle the same as the Greeks' gnomon method, but the level of difficulty is considerably higher.

Figure 3.5.9 shows how the Babylonians set out the dots (points) of the first four squares in a dot rectangle and by this means came to the formula:

$$1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{1}{3}(1 + 2n)(1 + 2 + 3 + \ldots + n) \qquad (4)$$

The figure shows that

$$1^2 + 2^2 + 3^2 + 4^2 = \frac{1}{3}(1 + 2 \cdot 4)(1 + 2 + 3 + 4) = 30.$$

3.5.7 *From Number Riddle to Algebra*

I ask the pupils in a seventh grade class to think of a number, double it, add 24 and then divide by 2. From the resulting number they subtract the original number. I then impress them by saying that I can "see" the final result ... It was 12, wasn't it? Did anyone get anything else? Oh well then, whose thinking has gone wrong, yours or mine? We repeat the exercise in similar variations. *Must* the answer in the game above always be 12? Let us try a little algebra: We let the original number be a. It is, of course, different for different people and so we cannot do more than give it a letter a, which could represent any number at all.

And now we write, step by step:

double the number:	we get	$2a$
add 24:	we get	$2a + 24$
divide by 2:	we get	$\frac{2a + 24}{2}$

Subtract the original number: the result is

$$\frac{2a + 24}{2} - a.$$

Can this expression be simplified? Yes!

$$\frac{2a + 24}{2} \text{ is quite simply } a + 12.$$

From this it follows $a + 12 - a = 12$ and the reading of minds is revealed for what it is!

In this manner one can exercise a class in algebraic simplification in an enjoyable way.

The step up to equations is now not especially large:

I am thinking of a number. I increase it by 8 and divide by 5. I get 7 as my result. What number was I thinking of?

One can solve this problem in the head, or preferably with pen and paper, and notice how the solution came about. It is both interesting and important that we now and then observe how we think.

It usually turns out that not all students (of those who actually solved the problem) can clearly account for *how* they solved the problem. But some students succeed in observing their thought process. They say, "First I took 7 times 5," since division gave 7. 35 must then have been the number before I divided by 5. Then I took away 8 and got 27, which is my answer.

As we see, the solution goes backwards relative to the statement of the problem:

Problem ↓	*Solution* ↑
Think of a number	The number was 27
increase by 8	before this it was 35 - 8 = 27
divide by 5	before this it was 5 · 7 = 35
result 7	the result was 7
What was the number?	What was the number?

Let us now look at another problem taken from everyday mathematics:

By adding salt to 5 kg of 4 percent salt solution, we wish to increase the concentration to 5 percent by weight. How much salt need be added?

In this problem we do not know in advance how much the salt will weigh, and therefore we cannot do a kind of "backward calculation" to find out the desired amount of salt. The problem is not so amenable to working in the head. For problems of this and similar types, equations are an effective tool. How do we go about setting up an equation?

We return to the riddle above and let x denote the unknown number we wish to calculate. Now we "translate" the problem wording step by step to an algebraic expression:

think of a number:	we call it	x
increase by 8:	we write	$x + 8$
divide by 5:	we write	$\frac{x+8}{5}$
the result is 7:	we can write	$\frac{x+8}{5} = 7$

We have thus come to two different expressions for the result:

$\frac{x+8}{5}$ on the left and simply the number 7 on the right.

We connect the two sides with an equality sign and have then an equation. Now the problem is to *solve* the equation

$$\frac{x+8}{5} = 7$$

We are to get it to the point x = some number.

In other words, we want to try to "get x by itself."

What does the equation tell us? $x + 8$ divided by 5 gives 7. We can multiply *both* sides by 5 and in that way maintain equality:

we get $$5 \cdot \frac{x+8}{5} = 5 \cdot 7$$

that is, $$x + 8 = 35.$$

This equation applies just as truly as the equation

$$x + 8 - 8 = 35 - 8$$

which is obtained by subtracting 8 from *both* sides.

We obtain the solution hereby: $x = 27$.

If we compare solving the equation with our solving of the riddle earlier, we find that we have taken the same steps. The advantage with the equation is that one can (with practice) write the equation down in the same order as the problem's wording and then "technically" go about solving for the desired value. We avoid the necessity of "thinking backwards."

From a *pedagogical* standpoint it is often better to activate students with more "figure-it-out-in-your-head" solution methods than to have them carry through mechanical solutions. But with a problem such as the salt problem the practical advantages of using an equation are considerable:

We let x kg be the amount of salt which the equation asks for.

Now we "translate" according to the text:

We have to begin with:	5 kg 4-percent solution
We add x kg salt:	The solution then weighs $(5 + x)$ kg
This solution has 5% salt by weight	5 % of $(5 + x)$ is $0.05\,(5 + x)$ kg salt (5)

Do we have any value to set equal to $0.05\,(5 + x)$?

This expression represents the amount of *salt*. Can this be written in another way? Yes. We look back to the original quantity of solution and note that it contained 4% salt, i.e. $0.04 \cdot 5$ kg or 0.2 kg salt.

We add x kg salt:

the amount of salt becomes $(x + 0.2)$ kg (6)

Both (5) and (6) give the amount of salt, so the equation is:

$$x + 0.2 = 0.05\,(5 + x)$$

Step by step in proper order we now get the following equivalent equations:

$$\begin{aligned} x + 0.2 &= 0.25 + 0.05x \\ -0.2 &\quad -0.20 \\ x &= 0.05 + 0.05x \text{ (note that } x \text{ means } 1.00x) \\ -0.05x &= \quad -0.05x \\ 0.95x &= 0.05 \end{aligned}$$

Finally we "get x all by itself" by dividing both sides by 0.95.

$$\frac{0.95x}{0.95} = \frac{0.05}{0.95}$$

$$x = \frac{1}{19} \approx 0.053$$

The answer is therefore: 0.053 kg or 53 g of salt needs to be added.

I have consciously chosen to present a problem leading to a relatively difficult equation. In the sixth or seventh grades in school one must begin with easier equation problems. The risk is then, however, that the problem is so easy to solve in the head that the students protest what they feel is making an easy thing difficult by doing it with equations. The thing here is to find a good middle-of-the-road problem. It is desirable for the class at an early stage to gain an appreciation that equations are a working tool in the solution of problems.

3.5.8 Exercises

1. A cylindrical pipe has a wall thickness of t, inner diameter d and outer diameter D. State some relations between these three quantities.

2. We look back to the end of Section 3.5.2: is the last triangle obtained there identical with Pascal's triangle?

3. Can four numbers a, b, c, and d form a magic square of 2 x 2 boxes, such that the rows, columns, and diagonals (with two numbers each) all give the same sum?

4. What general results can one come upon if one continues the study in Section 3.5.4?

5. Show that every prime number greater than 3 can be written in the form $6n + 1$ or $6n - 1$, where n is a natural number (i.e. a positive integer).

6. Show with the help of Exercise 5, that if p is a prime number greater than 3, then

$$p^2 + 2$$

cannot possibly be a prime number.

7. Try to prove by induction the beautiful formula

$$1^3 + 2^3 + 3^3 + \ldots + n^3 = (1 + 2 + 3 + \ldots + n)^2$$

$$n = 1, 2, 3 \ldots$$

(Refer to Section 3.5.6)

8. We have 1 kg brass, containing 60 % copper.
How much more copper must we melt down and add if we wish to bring the copper content up to 70%?

3.6 Judgment and Misjudgment

3.6.1 Big Balls and Little Balls

Time after time everyday life asks us to make judgments. It can concern the most widely differing things, from simple comparisons of length to difficult evaluations of quality.

We are very familiar with comparison of lengths and distance. It is not hard to judge , in Figure 3.6.1, how many times longer line AB is than line CD.

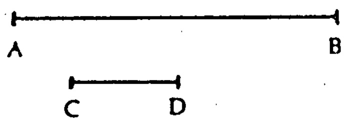


Figure 3.6.1

Much more difficult are judgments of distance "in depth," for example, across a bay or in the mountains. As far as time intervals are concerned, we all know how difficult it is to free ourselves from the subjectivity of our own experience. With two- and three-dimensional things we usually also have difficulty making reliable comparisons.

I don't know how many times I have begun a geometry lesson in the eighth grade with the following little story:

For a birthday party the hostess had made marzipan balls, solid, in two sizes. Some of the balls were twice as large in diameter as the others. The hostess offered them to her guests. The first guests each took a small ball; then someone took two small balls. The next guest thought, "I can just as well take a large ball instead of two small; there isn't much difference." After that another guest reasoned, "I'll take three small balls; I'm sure it isn't more than one large."

"What do you say, kids? How many small balls could one politely take, if one didn't wish to take more than the one who took a large ball? Can you estimate? You may very well — if you wish — use decimal fractions. Does a large ball contain just as much marzipan as 3 small, or 5 or 2.9 — or how many?"

Most students have guessed that 1 large equals 4 small. A few answered 3; a very few have answered with a decimal or with 5. Only seldom has someone answered 8, and then he was met with surprise and disparaging glances from his classmates.

Even in those classes which have been generous in their estimates, or where someone has answered 8, it has been advantageous not immediately to say if the answer is correct or incorrect. It is more fruitful to let the class itself judge the reasonableness and correctness of the answers.

I have therefore spun out the tale and told how the hostess also served jellied candies in the shape of solid cubes, some of them with double the edge width of the others. How many small cubes together would make as much candy as one large cube?

The answer "four" usually persists, but it comes a little doubtfully. And it doesn't take long before some students eagerly and with conviction in their voices say: "8 cubes." They are happy to go up to the blackboard. There they quickly sketch a cube, which, like a Christmas package, is divided into quarters on each side (Figure 3.6.2). There is no doubt about it: the large cube corresponds to 8 small cubes.

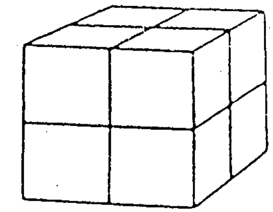


Figure 3.6.2

Everyone now suspects that the same holds true for the marzipan balls. According to the famous and classical "method of exhaustion" by the Greek Eudoxus (408-355B.C.), one can see that doubling the ball diameter gives an 8-fold enlargement in volume: one inscribes an endless series of smaller and smaller cubes within the ball, so that the total volume of all the cubes begins to approach the volume of the sphere as a limit.

In what context does a doubling of a length dimension result in a quadrupling? The class thinks this over, and there may come widely differing answers, which yet are basically correct. For example: if one doubles the radius of a cake but keeps the height unchanged, then the cake will be "4 times as big." Or: if one makes the side of a square twice as long, then the area will be four times as big.

In the continuation of our work, we eventually gain important knowledge about scaling. We set up a table:

When the length is	*... then the area is*	*... and the volume*
doubled	4 times as big	8 times as big
tripled	9 times as large	27 times as big
4 times as long	16 times as large	64 times as large
10 times as long	100 times as large	1000 times as big
half as long	1/4	1/8
one-third	1/9	1/27
one-fourth	1/16	1/64
1/10	1/100	1/1000

The numbers in the area column are the squares of the length numbers, the numbers in the volume column are the cubes of the length factors.

In summary: Area scale is the square of the length scaling.
Volume scale is the cube of the length scaling.

The most important application is without doubt converting between different units of measure for area and volume.

We have come up with an important rule, but we must time and again practice with what appear to be different examples:

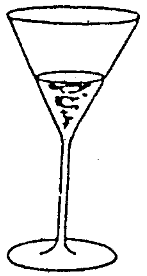


Figure 3.6.3

1. I sketch a cylindrical glass on the board: "Here is water up to this height. Now we fill water to double the height. How much water do we have in the glass now?" No one is fooled.

2. "Here is a conical glass (Fig. 3.6.3). It is filled up to half the height. If we now fill it up to the brim, how many times more water have we?"

A pupil remarks: "One can see that you have experience with this stuff!" After that encouragement the hands start going up. The first answer: four times as much. The second, third and fourth answers: four times as much. Are we all agreed on this? Doubt, until someone comes up with the right answer: 8 times as much. This time the pupils are surprised not so much over the answer as over the fact that they let themselves be fooled.

Repetition is the midwife to understanding, but it ought to be interesting and not just routine.

During a repetition one can readily tack on something new. Figure 3.6.4 is taken from Huff. It shows two sacks of money. The one has double the dimensions of the other. We suppose that the figure is meant to illustrate the doubling of profits in a business and thereby show how "successful" the company is. We might ask the class if they have any comments on the figure. Does it well illustrate the company's doubling of profits? The class need not think long before they see through the trick which lies behind the figure. What is it?

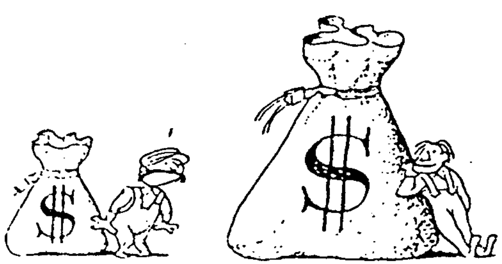


Figure 3.6.4 (From Huff, "How to Lie with Statistics.")

If the sack stands before us, how much more would you think the big one holds than the small one? When we see the sacks drawn three-dimensionally as in the figure, most people get an *unconscious* feeling that the big sack holds much more than twice the contents of the little one. Herein lies the trick. The volume ratio is, in reality, as we have seen, 8 to 1. More correct would be, for example, to illustrate the doubling of profits with a bar diagram (Fig. 3.6.5).

The three-dimensional drawing occurs quite often but sometimes the illustrator does not seem to have thought of the trick which it contains.

A problem concerning area and volume scaling, which occupied Galileo and can generate good interest in a ninth grade class, is taken up in Exercise 6.

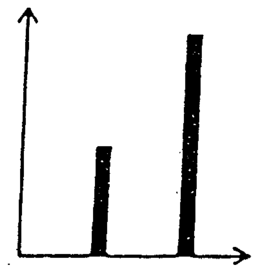


Figure 3.6.5

3.6.2 *Beware of Averages*

Among types of judgment and comparisons we make not least are averages. Oil consumption, food costs, speeds, etc. are all generally stated as averages. Suppose we are travelling by car, have covered a distance of 600 km. and been driving 8 hours including breaks. For fun we want to know what our average speed has been, including the breaks. This might even be of value with thought of planning for future trips by car. We find that our average sped was 600 km/8 hours, i.e. 75 km/hr.

Let us now consider the following problem: an airplane flies a certain distance at a speed of 800 km/hr. and immediately returns with a speed of 1000 km/hr. What average velocity has the plane kept? It is tempting to answer 900 km/hr. This is incorrect, however. If the distance were 10,000 km, for example, then the flying time would be

$$\frac{10{,}000}{800} + \frac{10{,}000}{1{,}000} = 12.5 + 10 = 22.5 \text{ hours.}$$

An airplane flying at 900 km/hr. the whole time would need the time

$$\frac{20{,}000}{900} \text{ hours, i.e. } \quad 22.222\ldots \text{ hours} \quad \text{to make the trip,}$$

which gives too short a flight time. The correct average speed must therefore be less than 900 km/hr. We see here how easy it is to get lost when calculating averages.

The airplane in our example returns at greater speed than during the outbound flight. The return thus goes more quickly, which means that the plane is travelling with speed 1000 km/hr for a shorter time than with speed 800 km/hr. If the plane had flown at these two speeds during equally long *times,* then the average speed would have been 900 km/hr, i.e. the average of 800 km/hr, since the distance covered in a given time is proportional to the speed.

3.6.3 *Per Cent Calculations*

Imagine how difficult it would be to get anything out of data and statistics, if we could not state the results as per cent!

Let us suppose that a person P owns three sevenths of the stocks in a company X and five elevenths in another company Y. In which company does P have the greater share?

It is not easy to compare directly $\frac{3}{7}$ and $\frac{5}{11}$. We often use per cent values and find that $\frac{3}{7}$ corresponds to 42.9% and $\frac{5}{11}$ to 45.5%. With that the comparison is completely clear.

In the old Swedish grade school the pupils were drilled in calculating how many per cent higher or lower a certain price was, compared to another. We need not long for a return to the old school methods, but considering how often percentage calculation is used in daily life, not the least in commercial advertising, it must be a part of general education to know a bit about such calculation.

If a firm wishes to sell out a lot of goods for half price, it may quite simply advertise: "These goods now selling at half price." Or: "50% discount on these goods."

But a more common way of advertising nowadays is:

"Get double value for your money." Or: "Get 100% more goods for your money."

All these expressions say the same thing, but doesn't it sound more tempting to the customer to get 100% more for his money than a 50% discount?

Per cent calculation is perhaps that area of everyday mathematics which offers us the most varied collection of problems, and on top of that, problems with the fresh taste of real life. Pupils become more aware of what it is one compares with, and they see how the results take on different expressions and have different values depending upon what one uses as a reference.

Per cent calculations also give problems which invite estimation with head calculation. Let us examine the declaration of contents of Falukorv (a favorite Swedish sausage): It states that the meat content is 68 % (by weight). The meat ingredients are then declared separately

beef 45%
pork 35%
and lard 20%

How much beef and how much lard, respectively, are contained in Falukorv?

Since the meat content is 68%, it represents about $\frac{2}{3}$ of the weight. The fraction of beef would then be $\frac{2}{3}$ of 45% or about 30%, while the fraction of lard would be $\frac{2}{3}$ of 20% or approximately 14%. One can find good examples in books like D. Huff's *How to Lie with Statistics*, but perhaps training in per cent becomes most alive when one picks out interesting articles from the daily newspaper. I reproduce here a few clippings to show what breadth such examples can have.

1. "Enormous mark-up! What does a farmer get for a kilogram of carrots at the wholesaler — and what does the housewife pay for the same carrots at the grocer's or market square? Express Patrol found that there is an enormous mark-up on certain vegetables and flowers. Carrots go up by 350% from wholesaler to retailer and broad beans by 600% during the same journey." (From an article in the Swedish newspaper *Expressen.*)

2. "1900-1977: Divorce up by 800%. Divorces in Sweden have increased markedly, looked at from a historical perspective. The increase is 800-900% since the beginning of the 20th century. Of all marriages entered, 25-30% are dissolved as a result of divorce." (Dagens Nyheter, Sweden)

3. In a 1973 article the ownership of the Swedish domestic airline Linjeflyg was described as complicated (Figure 3.6.6):

Owners of the Domestic Airline

DK = the Danish government
N = Norwegian government
S = Swedish government
Ö = Miscellaneous interests in DK, N, and S.
S INT LT = Swedish Intercontinental Air Traffic
S STORF = Swedish big industry
DANSKT LS = the Danish Airline Co.
NORSKT LS = the Norwegian Airline Co.
AEROTR = Swedish Aero-transport Inc.

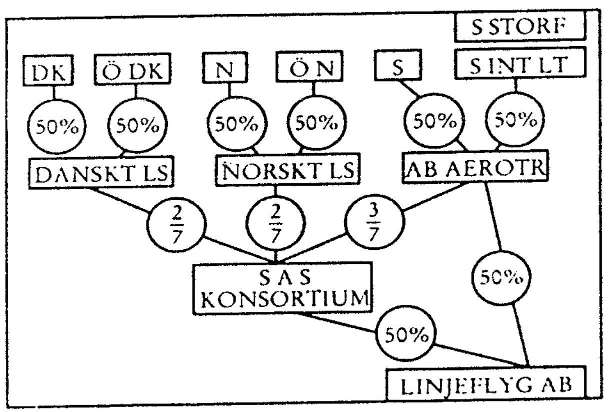


Figure 3.6.6

"Linjefyg's ownership is complicated. Danish and Norwegian interests, through SAS, have influence over the Swedish domestic airline, just as Swedish interests have in neighboring countries."

4. "New taxation Shocks" — newspaper headline following release of new property valuations in 1970.

"Nämdö community reports an increase of over 500% in taxation value for an area containing summer cottages. At last taxation the lots were valued at $1000. Now they are raised to $5000."

(Is the per cent figure correct?)

The numbers in these examples are a few years old, but we can find examples daily — equally varied and original — and use them in our work at school.

Opportunities for co-operation with colleagues in the natural sciences and particularly social studies abound. Inflation, wage increases, real increase in earnings, city planning, large scale production, etc., all give rich materials without further ado, materials which motivate the pupils in the higher levels in their studies.

3.6.4 Exercises

1. Refer to Example 3 above on Linjeflyg. What per cent of Linjeflyg is owned by Danes?

2. Refer to Example 4 above. Is the figure 500% correct?

3. The Boliden Company's Aitik mine is one of the "poorest" copper mines in the world: the ore is only 0.5% copper. How many tons of ore must be mined to get one ton of copper?

4. The price of a product is raised 25%. How big a discount can the store now give without the price going below the original level?

5. A community one year reduces its financial contribution to a school by 50%. The next year it raises it by 100%. (No other changes in the school monies are made). At what level, compared with the original year, is the school contribution now?

6. Galileo writes in his essays on "things which float on water":

> ... If one wishes to maintain the same proportions in the bones of the skeleton of a great giant, as we find in the ordinary human, then one must either use a stronger and harder material for the legs or accept a reduction in strength compared to common man, since if our height were to increase normally we would fall and be crushed by our own weight. Therefore, if a body's size is decreased, the strength of the body is not diminished in the same proportion; the smaller the body, the greater the relative strength. A small dog ought to be capable of bearing two or three dogs of the same size upon its back; but I believe that a horse could carry not even one of its own size.

Let us now suppose that adult mammal is 8 times heavier than a young animal of the same kind. How many times larger must the bones that bear weight in the adult then be, if we assume that the mechanical strength in both cases is in proportion to the cross-section of the bones?

7. A boat went 20 knots for 10 minutes and then 15 knots for 30 minutes. What is the boat's average speed during this 40 minute-period?

8. Calculate the average speed for the airplane of Section 3.6.2. It flies a given distance at 800 km/hr.. then turns around and flies immediately home at a speed of 1000 km/hr.

3.7 Nature's Geometry and Language of Form

3.7.1 Polyhedra

There is a rich literature on the many wonders of nature. Lovely color photographs show us examples of how nature has solved her problems in form and how she has given minerals, plants and animals color — not only beautiful but also functionally "correct." We have, perhaps, read some of these books or seen films, but how many observations have we ourselves made directly out in nature?

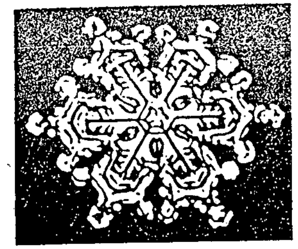
Quite certainly we have looked at plants and let their language of form and color fascinate us. What about minerals and particularly crystals? Possibly very few of us have sought out and scrutinized crystals in nature. Most often our interest is limited to precious jewels or indirect meetings through illustrations in books.

Experience in school has shown us that direct meetings in nature give pupils a much greater return and waken their enthusiasm considerably more, when they have had the opportunity to acquire knowledge of the subject in advance. (The same observation applies for subjects such as architecture, history, painting and other areas which we can meet up with during study trips and camps, for example..)

Let us begin with:

1. Crystal Shapes

Children get great joy from the first snow, sometime before Christmas, when King Bore sends his myriads of snowflakes. The earth



is covered in a white shroud, cleaner than most other materials which we can find. When the sun shines on the new snow, it often glimmers in different colors, and if we look carefully we can discover a little of the architecture which forms snowflakes.

The person who studied the snowflake's form probably more patiently than anyone else was the photographer W. A. Bentley, an American who developed a camera technique and photographed a vast number of snowflakes.

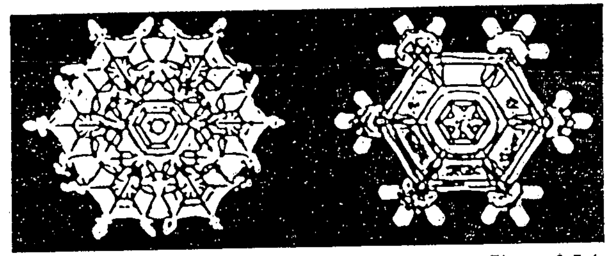


Figure 3.7.1

Together with W. J. Humphreys he gave out the book *Snow Crystals* (Dover Publishing, 1931) containing 2453 pictures of snow flakes — each picture different! One meets here an incredible multiplicity of variations on one and the same underlying theme: the architecture of the number six. Figure 3.7.1 shows some of Bentley's pictures.

Interest in crystals can be traced as far back as the historian's eye reaches. In Hellenic culture, Plato stands in the foreground among mathematicians' study of geometrical form. The huge mathematical work *Elements*, put together over 2000 years ago, comprised 13 volumes whose presentation reached its peak in the thirteenth and last volume, which treated regular three-dimensional figures, so-called regular polyhedra, or, as they also later came to be known, Platonic solids.

2. Regular Polyhedra

Polyhedra are solid bodies, bounded by plane surfaces: triangles, rectangles, and trapezoids, other polygons or a combination of such surfaces. (Poly = many, hedron = side, face, surface).

How many surfaces are necessary as a minimum, in order to make a polyhedron? Two is too few — we just get a "plough" shape. Is three enough? No, we only get a corner. But if we take four surfaces? Yes, we can construct, for example, a so-called tetrahedron with four triangles (tetra = four). We all have seen the "tetrapack" which milk, cream or juice comes packaged in. When the four triangles are equilateral and of the same size, we get a regular tetrahedron, and this we note down as one of the Platonic solids.

Do we know any of the other regular polyhedra? "The cube" is heard immediately from several directions in the room. Yes, it has six identical sides, six squares in a regular arrangement.

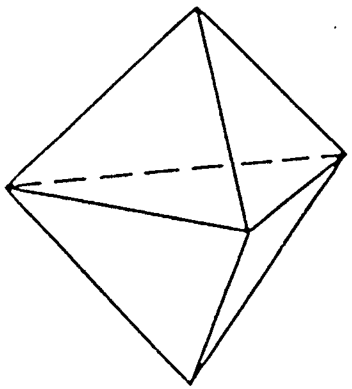


Figure 3.7.2

More examples? Someone says: "If we put two tetrahedra together so that two faces fit together over another, then we get a body which is formed by six equilateral triangles of the same size." This is correct, but ... How about the corners? "Are all the corners 'the same'?" — "What do you mean by that?" — "Well, if we look at the corners in the regular tetrahedron we notice that all the corners are the same, and this is also true for the cube. Does this apply to the double tetrahedron which we are now considering?"

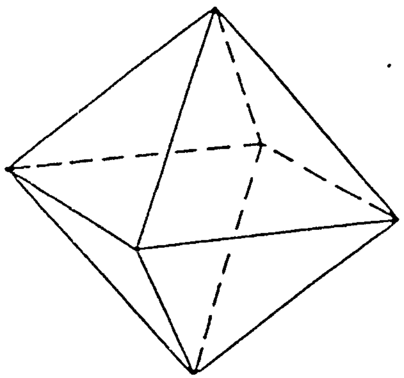


Figure 3.7.3

We look at a sketch on the blackboard. This time no one wishes to speak first. I must give a clue: look and see how many surfaces meet at each corner. Some now answer "three surfaces," others say "four triangles." Who is right? Proponents of these answers go up and point. Both are right in a way....

Let us look at Figure 3.7.2. Where are the corners with three triangles, where with four? At three of the corners, four triangles meet; at two corners (the top and bottom in the figure), only three meet. The corners are thus different. And we may therefore not call the double tetrahedron a regular polyhedron or a Platonic solid. In such a body it is not enough that the surfaces are all the same regular triangle or polygon. Even the corners must have the same form. We will therefore have to throw out our double tetrahedron.

What can we suggest now?

Someone says: "A double pyramid. We take two pyramids which have a square base and put the bases together. The walls are supposed to be equilateral triangles."

We get a solid bounded by 8 equilateral triangles (Fig. 3.7.3). How many meet in the corners? Around the "base square's corners" 4 triangles meet and at the top and bottom also 4. It looks promising. Can we make such a solid out of cardboard? In Figure 3.7.4 we see networks for cube and a tetrahedron. How would we draw a network for our double pyramid?

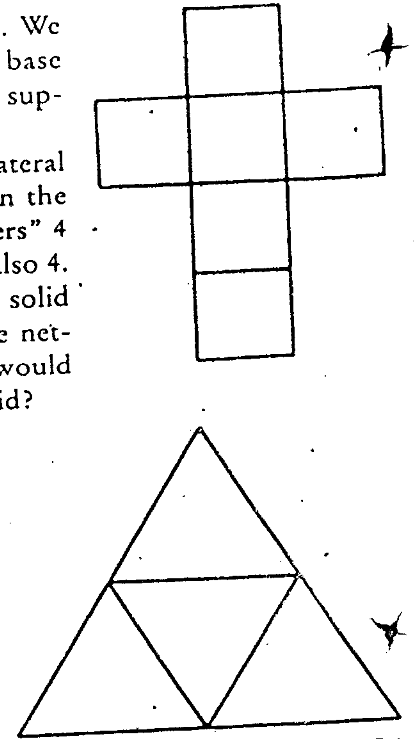


Figure 3.7.4

In Figure 3.7.5 we see one pupil's proposal, which ought to work well. We cut out the network and fold creases along certain edges. Finally we get the cellophane and the polyhedron stands finished. We twist it and turn it. Yes, all the corners are the same. No one doubts that this solid is entirely regular, and it becomes our third Platonic solid. It is called an octahedron, since it is bounded by eight surfaces.

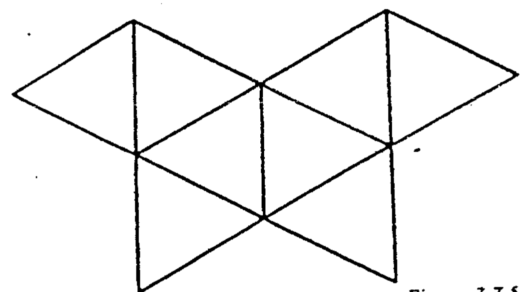


Figure 3.7.5

The class cannot come upon any further regular polyhedra — and it isn't easy. We therefore go on to comparing the three we have: How many corners, edges, and surfaces does each have?

We make up a table of the results:

	Corners	*Edges*	*Surfaces*
Cube	8	12	6
Octahedron	6	12	8
Tetrahedron	4	6	4

Someone discovers that the numbers 8 and 6 are just reversed in the cube and the octahedron. The number of edges is 12 in both. It seems as if these two might be related in one way or another.

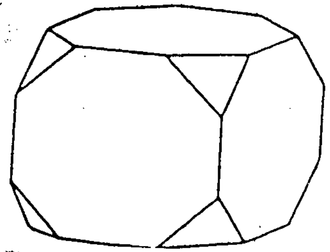


Figure 3.7.6

I now lead over into what appears to be a completely different question: What kind of solid arises if we gradually grind down the corners of a cube making the growing surfaces bigger and bigger?

We grind symmetrically so that equilateral triangles are created at the corners and so that all the corners are ground equally much. For example, what is the solid we get when the ground surfaces meet halfway along each edge of the cube?

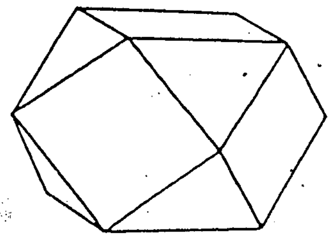


Figure 3.7.7

Figure 3.7.7 shows this stage and 3.7.6 a previous stage along the way, where each square of the cube has been transformed into a regular octagon (eight-sided figure).

What happens if we continue grinding further and go so far that the cube's square walls finally shrink down to nothing at their mid-points? Figure 3.7.9 shows the final solid. Figure 3.7.8 shows a transition stage (not so easy to draw!) where the ground surfaces consist of regular hexagons (six-sided figures).

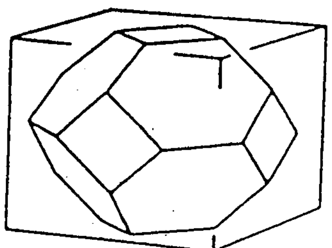


Figure 3.7.8

From the cube arose an octahedron, whose 6 corners are the mid-points of the cube's 6 squares, and whose 8 triangular faces came from the 8 corners of the cube. We understand now the relationship between the cube and the octahedron in the table. We are in addition completely clear over the fact that the octahedron is regular, since the grinding was done symmetrically.

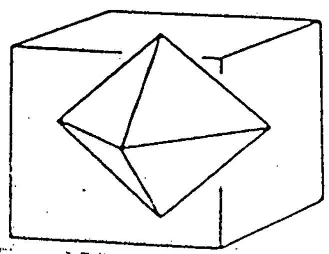


Figure 3.7.9

Someone wonders: what do we get if we grind down corners of an octahedron? Perhaps a new Platonic solid will emerge?

What happens? Let us sand down the corners in our imaginations, so far that only the mid-points of the eight triangular surfaces are left.

We are going to get a solid with 8 corners. A few pupils talented in geometry seem sure of themselves. (It is too bad when they say at once what they have found; it takes away their classmates' joy of discovery.)

It turns out that the cube emerges as the final form! We might have expected that, with a clue from the table. Halfway through the transformation we get the same solid as along the way from cube to octahedron, the solid in Figure 3.7.7 consisting of 8 triangles and 6 squares. It is called naturally enough a cubo-octahedron.

A new question: what arises if we grind down the corners of the tetrahedron so that only the mid-points of the triangular faces are left?

This question is easily answered: we get a new tetrahedron, turned upside down to the original.

In order to proceed we begin to systematize the Platonic solids in a table:

Cube:	6 squares, 3 squares per corner;	6 faces, 12 edges
Tetrahedron:	4 triangles, 3 triangles per corner;	4 faces, 6 edges
Octahedron:	8 triangles, 4 triangles per corner;	8 faces, 12 edges

What ought to come next? A solid with 5 triangles per corner? Or one with 4 squares per corner? "No, that is impossible. Then we couldn't fold up the cardboard walls." "Six triangles per corner won't work either; they fill out the whole cardboard surface around the corner." "But five triangles — that might just work!"

"What might the network for that solid look like?" I wonder aloud. Some suggestions come up. The starting point is that we draw two rows of five triangles. But how should these five-sided pyramids be put together? We help each other reach the conclusion that a "zig-zag girdle" of 10 triangles is necessary. Now proposals for networks start coming up: they consist of 5 + 5 + 10 = 20 triangles. (See Figure 3.7.10).

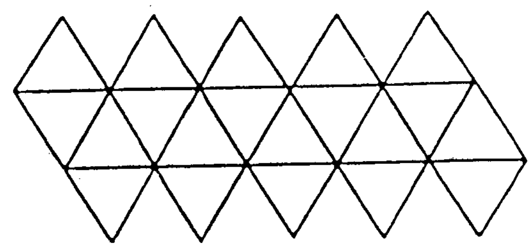


Figure 3.7.10

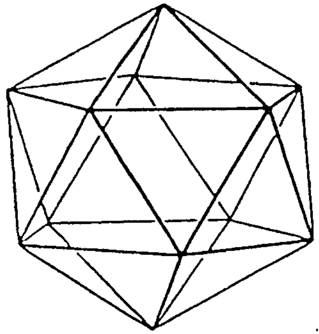


Figure 3.7.11

Scissors go into action: certain edges are scribed so that the faces can more easily be folded up. The pupils help each other in pairs with the tape.

Look, it fits together; it makes a roundish sort of solid!" It is called an icosahedron (icos = 20). It would be going too far to guide the class through an "existence proof" for this solid. For the majority, the successful taping together of the network is "proof" enough, and much more tangible than any theoretical arguments — at the 15 year-old level. (Figure 3.7.11)

We add to the table:

Icosahedron: 20 triangles, 5 per corner; 20 faces, 30 edges, 12 corners

What happens if the icosahedron is ground down, so that only the mid-points of the triangular faces are left? "It *has* to be solid with 20 corners and 12 faces."

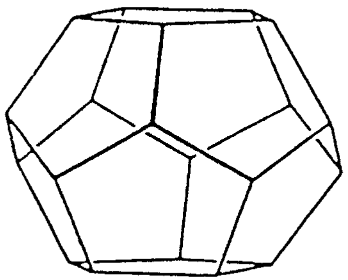


Figure 3.7.12

But what kind of faces? We think this over. How many corners must each face have? Its corners must lie within the five triangles which form each corner of the icosahedron. This means that the new solid's faces have five corners. Why yes, they must be regular five-sided polygons, so arranged that three meet in each corner.

The new solid is a sibling to the icosahedron in the same way that the cube and the octahedron are siblings to each other.

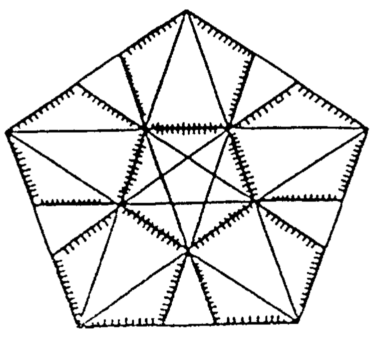


Figure 3.7.13

This fifth Platonic solid is called the regular dodecahedron, or more exactly, the regular pentagonal dodecahedron (Figures 3.7.12 and 13.)

Can a solid be constructed with 4 pentagons meeting in each corner? The answer is no, since four pentagons drawn

on cardboard must overlap each other. Can we succeed with a solid of three hexagons? No, they fill up the surface of the paper around a corner exactly.

We now realize that we have found all of the regular polyhedra which exist; that is, five.

Let us once again look at the pairs of opposites, the cube-octahedron and the icosahedron-dodecahedron:

Cube:	3 *four*-sided faces per corner; 6 faces, 12 edges, 8 corners
Octahedron:	4 *three*-sided faces per corner; 8 faces, 12 edges, 6 corners
Icosahedron:	5 *three*-sided faces/corner; 20 faces,30 edges,12 corners
Dodecahedron:	3 *five*-sided faces/corner; 12 faces, 30 edges, 20 corners

We say that these paired solids are duals of each other, two by two. The tetrahedron, which when grounded down, is transferred into itself, is said to be self-dual:

Tetrahedron: 3 three-sided faces/corner; 4 faces, 6 edges, 4 corners

3. Crystals — once more

Now, what do these five Platonic solids have to do with crystal forms? Are there crystals which take on the forms of Platonic solids? Yes. The very simplest example is kitchen salt. It can crystalize in cubic form. It can also have one side longer than the others and become box-shaped. The kitchen salt crystal is completely transparent and as colorless as ice. One might even mistake it for a piece of cut and ground ice. "Who ground it?" someone asks. "Nature herself — but she has not actually *ground* it; it *grew* that way, by itself!"

We take a look at gold-glimmering, cube-shaped crystals. On some of them sits, slightly crooked, a much smaller cube, "like a little crystal child," someone adds. "Is it gold?" — "No, it is pyrite, which is formed when iron and sulfur combine with each other in certain proportions."

I bring out some other gold-shimmering crystals, octahedra this time. Yes, these are also found in magnetite or lodestone (magnetic iron oxide, mined in Kiruna, Sweden, among other places) and in fluorite (calcium fluoride). The latter forms yellow, blue, violet, or yellow-violet octahedra. An almost pure regular tetrahedron form is found in Fahlband (German: Fahlerz); the crystals have a metallic glance and contain copper and sulfur, as well as one or several of the elements iron, antimony, arsenic, silver, and mercury (Figure 3.1.14).

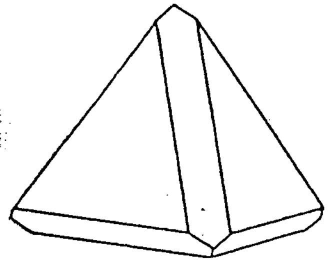


Figure 3.7.14

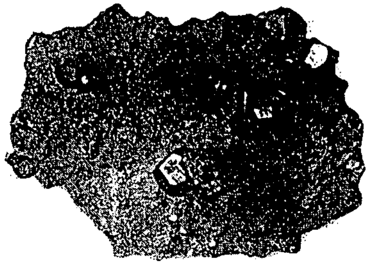


Figure 3.7.15

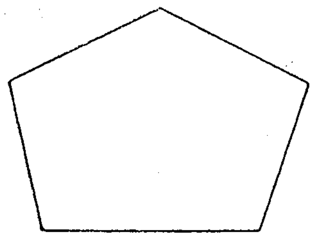


Figure 3.7.16

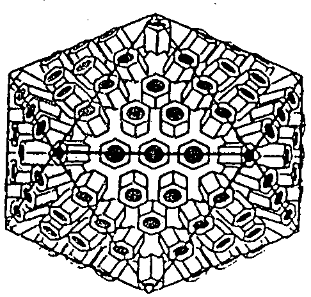


Figure 3.7.17
Schematic picture of a Herpes virus

Pyrite can also form dodecahedra. In Figure 3.7.15 we see a group of pyrite crystals in which nature has formed partial dodecahedra. If one looks very closely at such crystals, one should find that they are not completely regular. The five-sided figures which make up the faces of the crystal are a little bit too wide in one direction (Fig. 3.7.16).

4. Other Areas of Nature

Strangely enough nature does not produce crystallic *regular* icosahedra or dodecahedra. (We will not go further into this here). But in another area of nature, an area which could be said to lie between the mineral and plant kingdoms, there do exist regular icosahedra and in some case dodecahedra: among the viruses. Figure 3.7.17 shows a schematic picture of a Herpes virus, drawn according to pictures taken with an electron microscope. In Figure 3.7.18, also taken from *Scientific American* (Number 1, 1963), may be seen "shadows" of an insect virus, obtained by irradiating it in an electron microscope. The shape of the shadows reveals that the virus body itself has the form of a regular icosahedron!

Even radiolarians (an order of single-celled sea-animals with long slender pseudopodia "radiating" outward) show the regular icosahedron shape — namely in their skeletons, built of silica. They were studied by Ernst Haeckel, who made many interesting and detailed drawings of them (Fig. 3.7.19).

Numbers 2, 3, and 5 in the figure have the forms octahedron, icosahedron

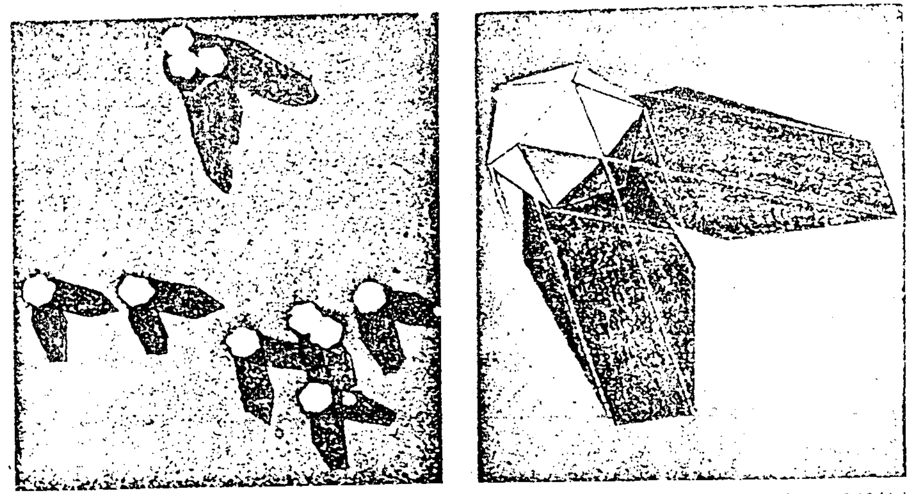


Figure 3.7.18 (From Scientific American 1963/1.)

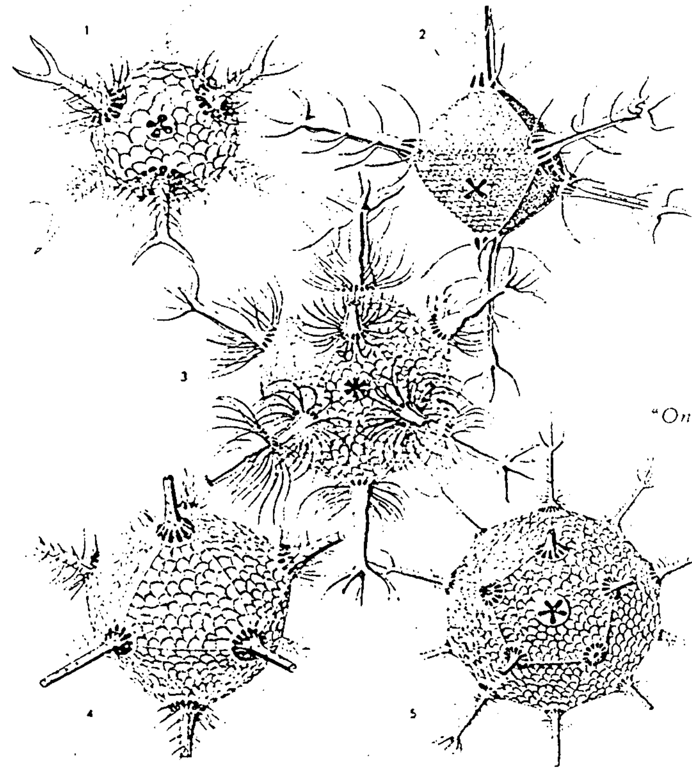


Figure 3.7.19
Radiolarian drawings
From d'Arcy Thompson
"On Growth and Form"

and dodecahedron respectively, and take their Latin names from these shapes: Circoporus octahedrus, Circogonia icosahedra and Circorrhegma dodecahedra. So here we have even been able to find examples of dodecahedra in nature.

When we say, for example, that fluorite crystals are octahedra, we must keep in mind the fact that the crystal-faces are not perfectly flat surfaces. The same applies to he radiolarians. But it is nonetheless obvious that the architectural form is regular. With the radiolarians the surfaces are somewhat curved, but the vertices form the corner-points of the Platonic solids. That pyrite crystals are not completely regular dodecahedra, however, has other explanations.

In *Scientific American*, Jan. 1983, there is an article "Platonic Chemistry" (p.59), where the synthesis of a molecular dodecahedron is reported. It was achieved by researchers at Ohio State University. In 1982, according to a Swedish newspaper, an Israeli physicist, Dan Schechtman, then working at the American National Bureau of Standards had succeeded in developing a technique, by which he manufactured dodecahedric crystals containing aluminum and iron atoms. The atoms took the positions of the twelve corners of a regular octahedron.

3.7.2 Curves and Curve Families

In spite of the richness in variety we would have to say that crystal forms are characterized by very strict "laws." For example, two surfaces always meet in a straight line edge. In plants and animals we find completely different principles of form. There the curve, the curling form, plays a major role. We shall soon pick out a few important examples, but let us first begin with the straight line and the swerving curve as basic elements for two-dimensional forms. Among curves we can consider the circle as the primary opposite of the straight line. This reflects itself in the tools we use: straight edge and compass, tools which trace their roots to Euclidean geometry's childhood 2500 years ago. If we wish to see curves as the result of one or more *movements*, we can begin with two basic motions: movement along a straight line and rotation.

Example 1: We can ask ourselves: what sort of curves arise when we combine the straight-line movement with a rotation? We let a point

simultaneously surrender itself to both a straight-line and a turning movement. What kind of motion does the point describe?

We can imagine that a little ball moves straight out from the center of a phonograph turntable, at the same time as the turntable rotates. How does the ball's motion look relative to the stationary table? It is a spiraling curve.

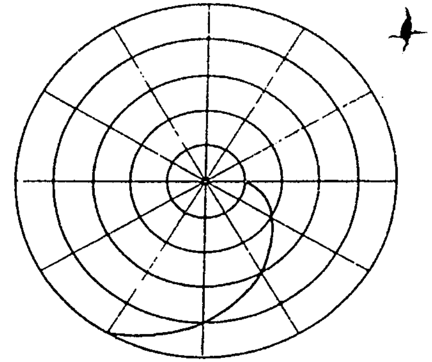


Figure 3.7.20

But we need to be more precise about how the movements come about. To begin with we choose the simplest motion: the point moves with constant velocity of 1 cm/second on the disc and the disc rotates at a constant speed of, let us say, 1/12 revolution per second around its center.

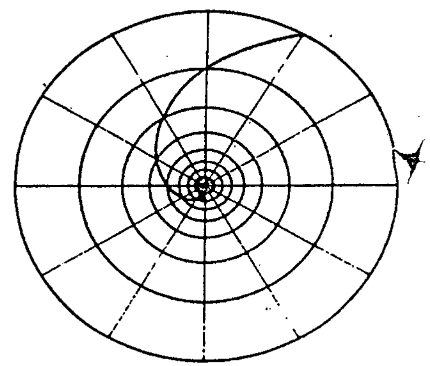


Figure 3.7.21

We can now draw a number of concentric circles a cm apart, and divide these into sections with 12 rays from the center. If the moving point starts at the center, it will describe the type of spiral motion shown in Figure 3.7.20. Irrespective of what values we give to the straight-line speed and the speed of rotation, just as long as they are constant, this kind of spiral is called an Archimedean or arithmetic spiral.

We now vary the theme and let the straight-line movement be accelerated, such that the moving point's distance from the center grows second by second according to the geometric sequence 4, 6, 9, ... (each radius = 1.5 times the next smaller radius). The

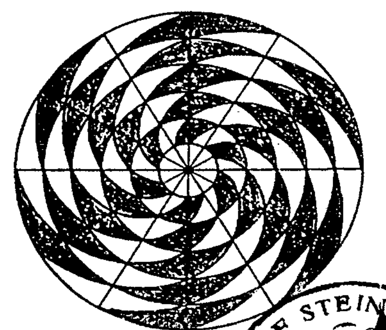


Figure 3.7.22

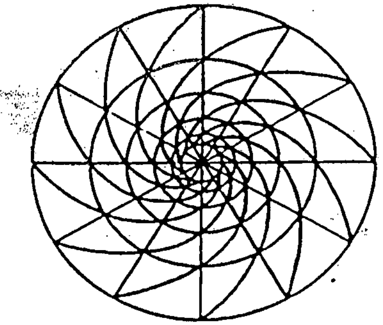


Figure 3.7.23

Figure 3.7.24
(From Adams/Whicher, The Plant between Sun and Earth, *1952).*

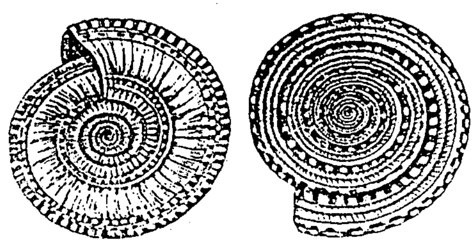
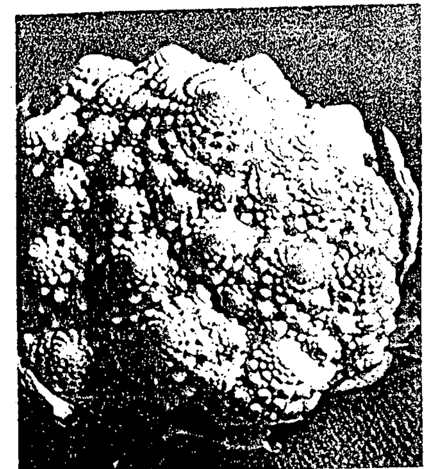


Figure 3.7.25
(From d'Arcy Thompson, On Growth and Form*)*

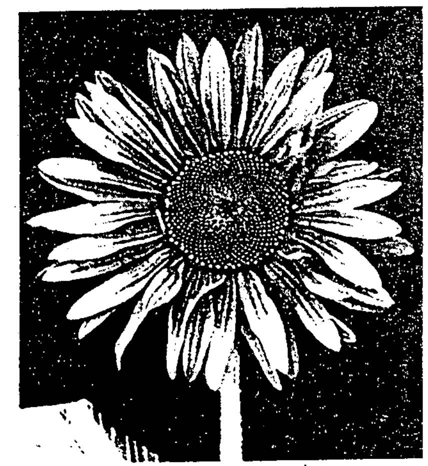
rotation is the same as before. We get now a considerably more dynamic spiral (Figure 3.7.21), the so-called logarithmic spiral.

Figures 3.7.22-23 show families of the Archimedean and the logarithmic spirals respectively. Diagonally situated "parallelograms" are colored in. What happens with these parallelograms, in their respective figures, as we move outward from the center? The class has no difficulty seeing how the Archimedean parallelograms change shape, they become wider, while the logarithmic spiral's parallelograms seem to keep their shape. They only get bigger. This observation is correct: one can show that the logarithmic spiral continually curves at the same rate, while the Archimedean spiral, as we have seen in the pictures, adapts itself more and more to the circle's form. Do these spirals exist in nature? Let us look at Figures 3.7.24 and 3.7.25 which show respectively the elegantly formed sea conch Nautilus (in cross section) and the mollusk Solarium perspectivus.

Figure 3.7.26 shows Italian cauliflower — an architectural masterpiece, where the spirals in the head of the cauliflower are found again in successively smaller scale in the miniature "heads" which make up the spirals. Here we have spirals of spirals of spirals, as far as the eye can reach. In this example and the next the spirals are logarithmic.



Cauliflower *Figure 3.7.26*



Sunflower *Figure 3.7.27a*

In Figure 3.7.27a we see a photograph of a sunflower (from Adams/Whicher, *The Plant between Sun and Earth*, 1952). Adams points to the sunflower's center, the glomerule or seed cluster, as an example of spiral formation in the plant kingdom. Let us study the photograph closely and notice that the cluster contains *two* systems of spirals: on the one the spirals turn clockwise in toward the center, in the other counter-clockwise. The latter spirals are longer.

Take a magnifying glass and try to count how many spirals there are, going clockwise and counter-clockwise respectively! We have counted spirals in sunflowers growing on the school grounds and come to the same result: 55 clockwise spirals and 34 counter-clockwise! Let us now reflect upon the following quotation from *Scientific American*, 3/ 1969:

> The most striking appearance of Fibonacci numbers in plants is in the spiral arrangement of seeds on the face of certain varieties of sunflower. There are two logarithmic spirals, one set turning clockwise, the other counterclockwise, as indicated in the illustration on the next page. (Figure 3.7.27b)
>
> The number of spirals in the two sets are different and tend to be consecutive Fibonacci numbers. Sunflowers of average size usually have 34 and 55 spirals, but giant sunflowers have been developed that

go as high as 89 and 144. In the letters department of *The Scientific Monthly* (November, 1951) Daniel T. O'Connell, a geologist at City College of the City of New York, and his wife reported having found on their Vermont farm one mammoth sunflower with 144 and 233 spirals!

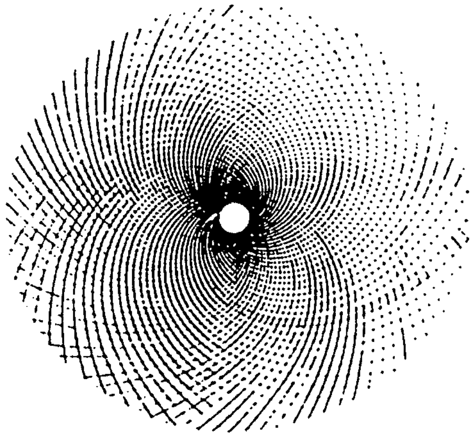


Figure 3.7.27b

Example 2: In Figure 3.7.28, points A and B are the points of origin for two motions (each represented by a family of circles) expanding outwards at a constant rate. What is the motion of a point which:

a) goes outward from A and simultaneously inward toward B?

b) removes itself simultaneously from A and B?

We first have to extend the families of circles so that they intersect one another. Then it is only a matter of drawing beautiful curves through appropriate intersections, and we get the family of curves in Figure 3.7.29:

the one family (a) shows us ellipses,

the other (b) portrays arcs off hyperbolas.

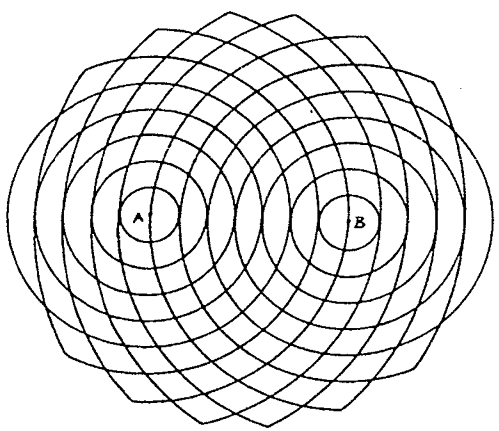


Figure 3.7.28

Such a pattern occurs when circles spread out in water from two centers; one speaks of interference patterns. Such patterns play a very important role in all technology which utilizes wave motion.

Example 3: Two straight-line motions can interact to create a curve in a completely different manner than they did in Example 2. We let two points A and B move along two respective lines a and b with, let us say, the same constant speed. We mark off the position of the points at evenly spaced intervals of time (Figure 3.7.30).

How do the connecting lines move? We have only to connect successive A-points with their respective B-points and see what emerges: Figure 3.7.31. The line rotates and creates a beautiful curve, a close relative to the ellipse and the hyperbola, namely a parabola. (The figure shows only an arc of the parabola.) The interesting thing is that the figure obtained gives the impression of three-dimensionality: one can see a saddle-like surface. In actual fact the figure may be seen as the plane projection of a so-called parabolic hyperboloid, a saddle-shaped surface with many beautiful characteristics.

The main importance of Example 3 is, however, that the pupils experience a completely different way of generating a curve than

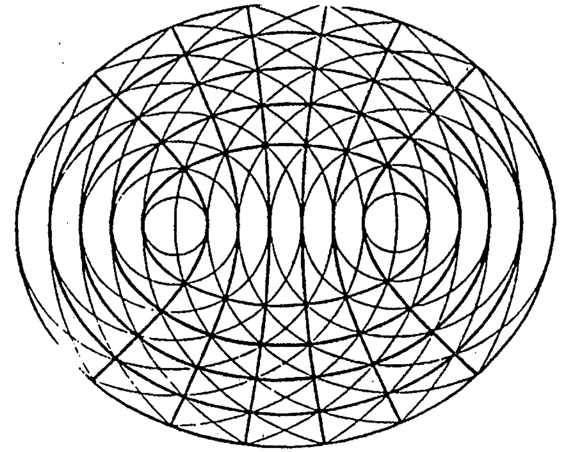


Figure 3.7.29

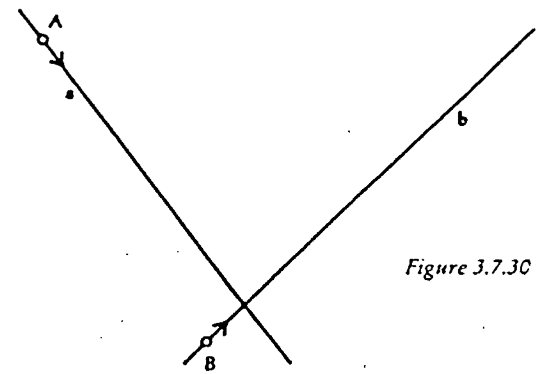
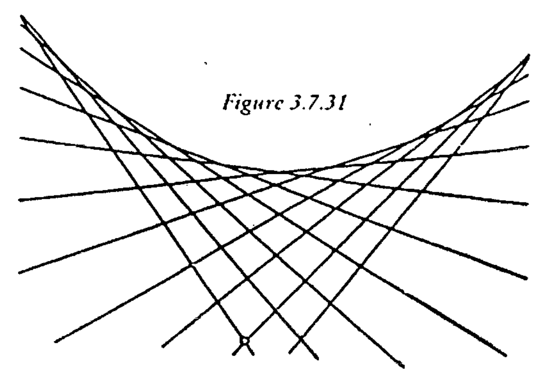


Figure 3.7.30

Figure 3.7.31



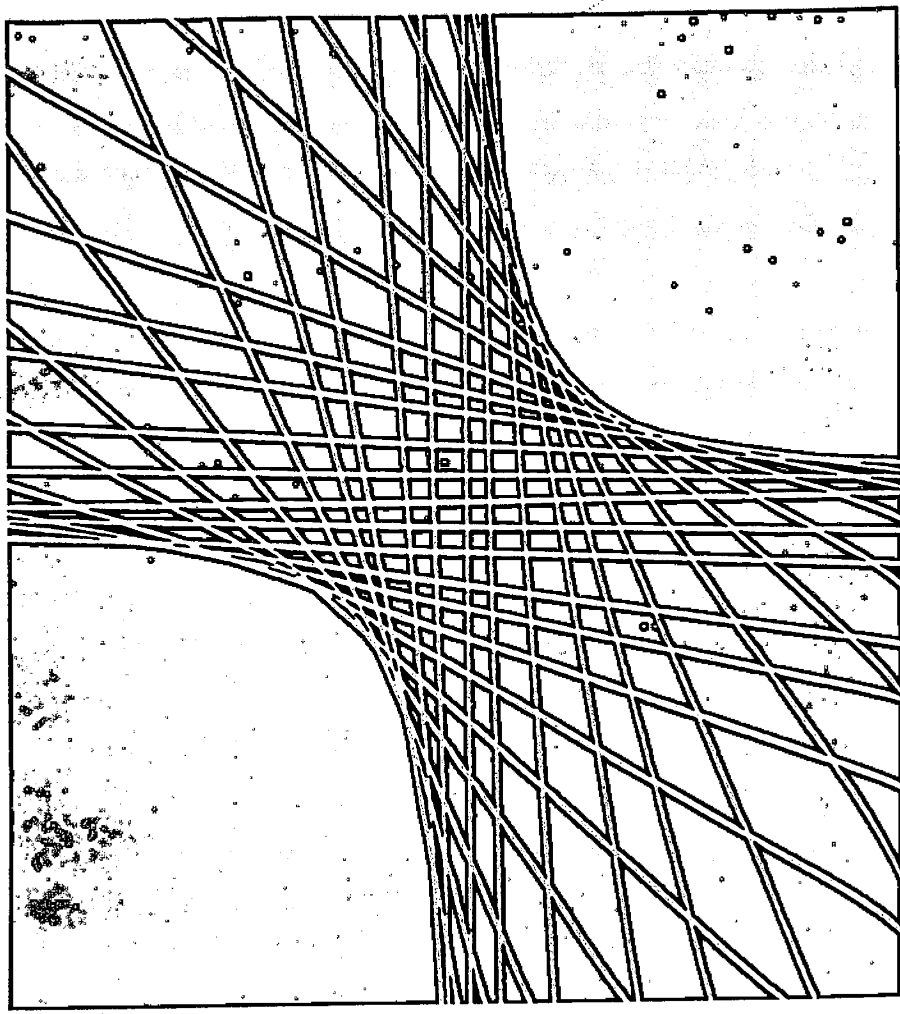


Figure 3.7.32

the otherwise common method — that which we used in Examples 1 and 2: constructing points on the curve. In Example 3 the curve is *shaped* by straight lines which will become tangents to the curve. There are many beautiful examples of constructing curves as the boundary of a group of lines or planes, so-called envelopes. Such figures are especially beautiful if they are drawn with white pencil or ink on black or blue paper. Figure 3.7.32 shows an example.

3.7.3 *Exercises*

1. Figure 3.7.4 shows networks for the cube and the regular tetrahedron. How many *different* networks exist for these respective solids? Two networks are considered identical if the one can cover the other. Rotation and mirror imaging thus do not give new networks. In addition, adjacent faces in the network must have a common edge (it is not sufficient for them to have only a common corner).

   Finding *all* possible different networks for the cube is a combinatoric-geometric problem which usually really gets the students engaged (ninth grade). One of the questions which comes up is, "How do we know when we have actually found *all possible* networks?" We have to figure out some sort of system for numbering the networks. Yes, we have to order them in a logical sequence. This is a task which fits well together with the problem we discussed in Section 3.2.

2. Make a regular dodecahedron and study it to see if you can find 8 points which would form the corners of an "inscribed" cube.

3. Draw a sketch showing how four of the corners of a cube determine a regular tetrahedron.

4. Start with the same drawing as in Exercise 3 and continue by drawing in the tetrahedron which is determined by the other four corners of the cube. Finally, try to make clear the solid which these two tetrahedra together form (Kepler's twin tetrahedra).

5. Draw networks for the solids shown in Figures 3.7.6 and 3.7.8, and do this in such a way that these solids may be inscribed in a cube of chosen size.

3.8 Curve Transformations

Transformation is one of the most important principles of nature's forms. Perhaps the most striking is the butterfly's development from the larva by way of the pupa stage. Other great transformations can be found, for example, in the tadpole's growth into frog and, in human beings, in the embryo's transformation into fetus. Even plants give us beautiful examples. Take, for example, the different stages of the dandelion. Simply observing the changing form of the flower central base, we can discover a surprising sequence of phases. The central base transformation makes possible the changes in the various stages: bud, flower, seed bud, seed sphere (Figure 3.8.1).

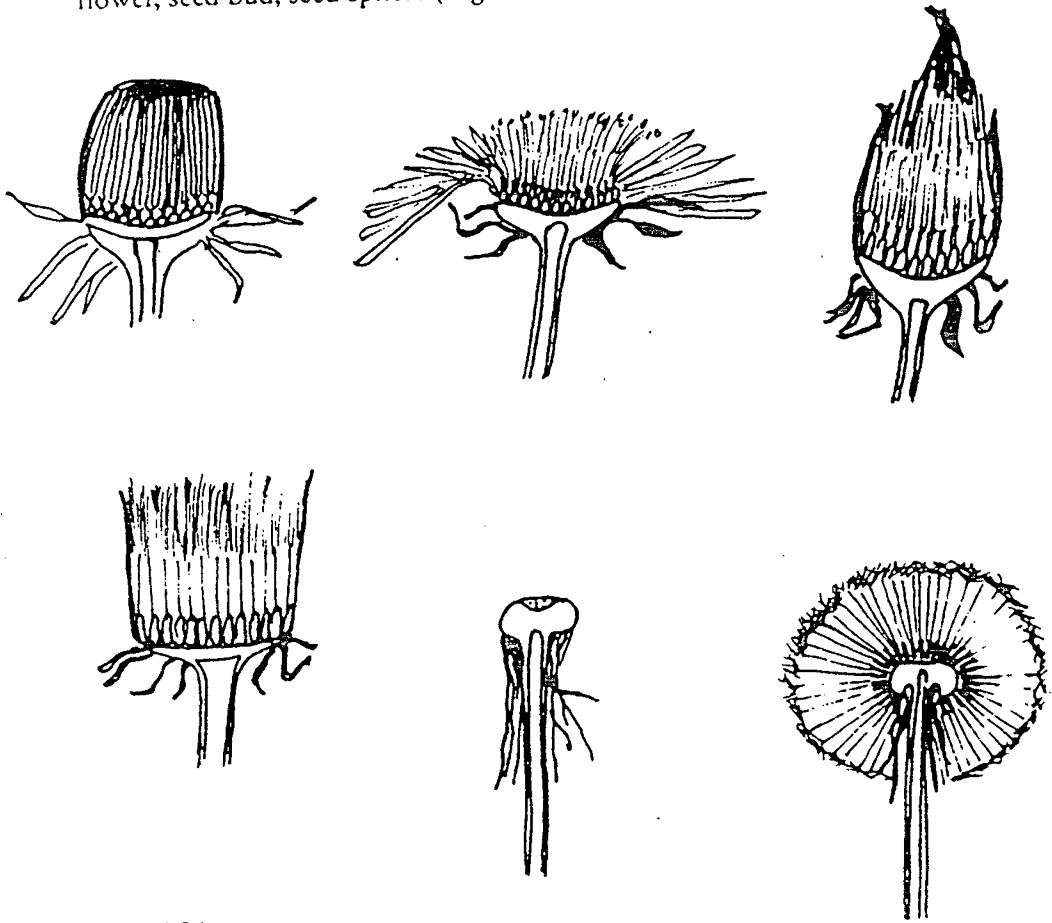


Figure 3.8.1

We will, by and large, here have to limit ourselves to the simpler type of transformation, which can be called curve transformation. So-

called projective geometry, which in this book can only be hinted at, lays the foundation for more sweeping metamorphoses.

Next we shall look at four examples, which do not require special prerequisite knowledge.

3.8.1 Four Examples

Example 1: From classical Greek geometry we take the curve which is called cochoid. (It was introduced by Nichomedes as a tool for solving the classical problems of doubling the cube and trisecting the angle.) A line a and a point P (not on the line) are the starting elements. Through P a line p is rotated. We call the intersection with line a X. From this intersection X we now mark off line p a given distance XY, in the direction of point P. To begin with we take XY as a small distance. What curve is described by the point Y, as line p rotates about point P? And in particular: how does the curve change as the length XY grows?

(Nichomedes took interest primarily in the case where the distance XY is marked off so that Y lies on the opposite side of P from line a.)

Point P can be suitably placed at 50 mm away from line a. A few symmetrically chosen lines through point P can then represent the family of lines p. For the dis-

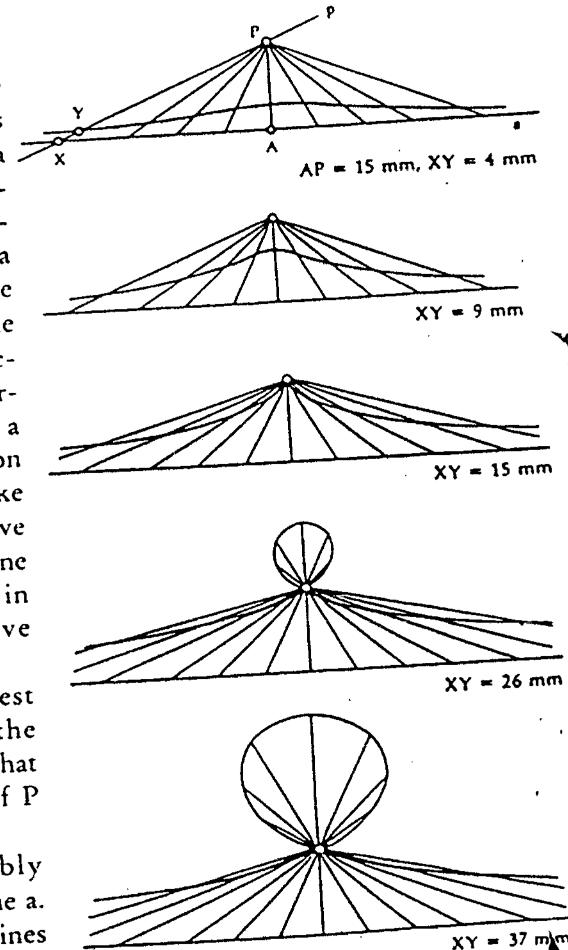


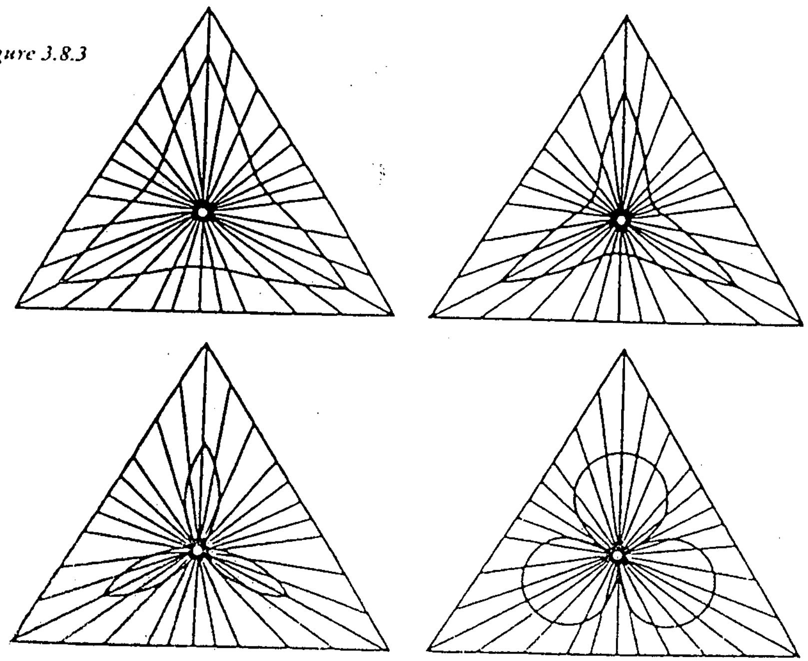
Figure 3.8.2

tance XY we might choose the following: 15 mm, 35 mm, 50 mm, 60 m and 80 mm. Each of these values corresponds to a conchoid curve (se Figure 3.8.2). (Other values have been chosen for the lengths XY and A in the figure, with consideration to the size of the book.)

We may now in our imaginations follow the conchoid's continuo transformation as we increase the length XY. In the vicinity of point arises a little "hill" which changes into a pointed peak, when XY = AP When XY becomes larger than AP, the conchoid makes a loop above P; becomes a so-called double point. After that the loop grows withou limit as XY is allowed to grow beyond all bounds. To the sides awa from point P the conchoid comes nearer and nearer to line a and arbitrar ily close to the line on both the far left and the far right. Line a is th asymptote for all the curves in the family.

If instead of drawing conchoids with respect to a single line, w draw conchoid arcs in relation to three line segments forming an equilat eral triangle with point P at its center, we obtain the closed curves a shown in Figure 3.8.3. The transformation then makes an even stronge impression.

Figure 3.8.3



Example 2: We are given a circle C and a line passing through the circle center M. On a line a we are also given a point F, which coincides with M to begin with. What does a curve K look like, whose points lie equally distant from circle C and point F? (F and M are the same, to start with.)

A moment's reflection shows it to be a circle with center at F, concentric to C and with half the radius of C.

We now let C grow by moving M to the right on line a, keeping the left-hand intersection of the circle with line a fixed at point V. We get a family of larger and larger C-circles, all passing through V. How does the curve K change as C grows? (We note that V is fixed.)

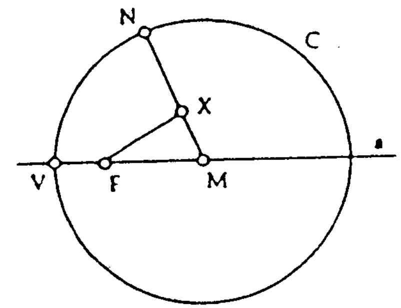


Figure 3.8.4

The pupils make free-hand sketches, before we begin to analyze the problem more closely. While they are drawing their figures, they come upon the fact that K must be a closed curve. Is it oval, egg-shaped, or what? In Figure 3.8.4 X is a variable point on curve K. We extend the line segment MX out to intersect the circle at N, obtaining the length MN. In order for X to lie as far from F as from the circle's edge we must have FX = XN. From this it follows that

FX + XM = NX + XM = NM = C's radius FX + XM is thus constant for all points X on curve K. The curve is determined by the equation

$$FX + XM = \text{constant} \; (= \text{radius}) \qquad (1)$$

From (1) it follows further that the points F and M play identical roles for the curve; it must be symmetrical not only about the line a but also around the line bisecting FM: the perpendicular bisector of FM. The curve is thus doubly symmetrical. Equation (1) specifies, quite simply, the curve which we call an *ellipse* (Fig. 3.8.5).

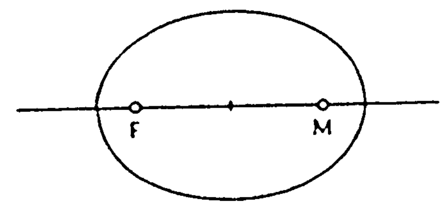


Figure 3.8.5

How does the ellipse change as the circle continues to grow? A few constructions show that the ellipse not only becomes larger but also more drawn out. Finally, if we imagine a circle C with an

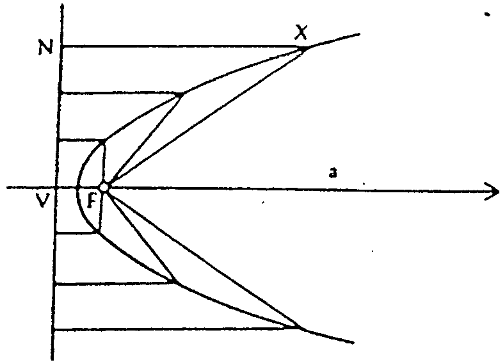


Figure 3.8.6
Parabola. FX = XN

"infinitely large radius", then M must be "infinitely far away," and of the circle we see only a straight line passing through V at right angles to a. Our curve is then stretched out infinitely and is called a parabola (Figure 3.8.6).

What happens with our curve if the straight-line circle "flips over" and becomes a circle with center to the left of V?

Once again we have that XF = XN, but this equality can now be written XF = XM - r (Figure 3.8.7) or XM - XF = r, where r = the circle radius (2)

X now describes half of a *hyperbola.*

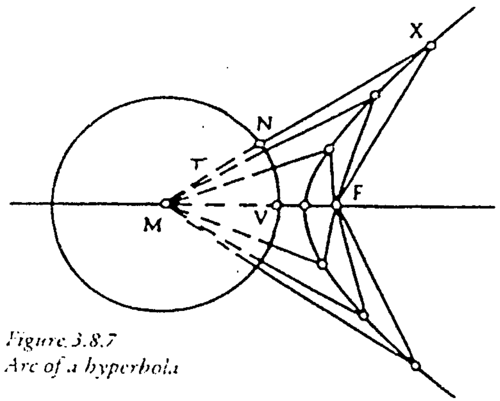


Figure 3.8.7
Arc of a hyperbola

If we change equation (2) to XM - XF = ± r
that is |XM - XF| = constant (3)

then the points M and F once again play equal roles, and we have the complete hyberbola (Figure 3.8.8).

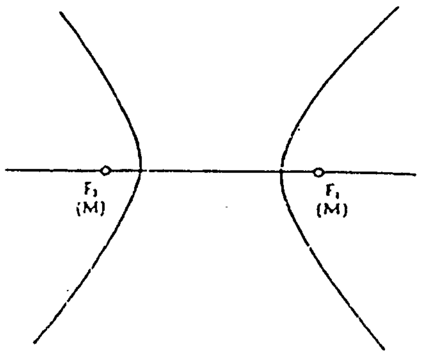


Figure 3.8.8
Hyperbola

As the value of the constant in (3) goes toward zero, the hyperbola straightens itself out; its two branches come nearer each other and close in from both sides on the line bisecting and normal to the segment MF.

If the circle C we start with is very small, we can

summarize the curve trnasformation as going from a little circle through growing ellipses to an infinite parabola to flatter and flatter hyperbolas and finally to a double line. (Figure 3.8.9.)

The equations
(1) XM + XF = constant and
(2) |XM - XF| = constant
show us that ellipses and hyperbolas can be seen as addition and subtraction curves, respectively. This characteristic of the ellipse and the hyperbola as pictorial representations of addition and subtraction, respectively, is brought out clearly in the curve families (ellipses and hyperbolas) which we constructed earlier in Example 2 of the previous section 3.7.2.

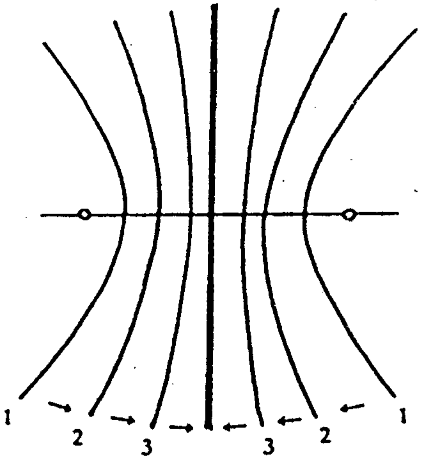


Figure 3.8.9

Example 3: What curves might be said to correspond to multiplication and division, respectively?

If A and B are two given fixed points and u and v are the respective distances to them from a point X, then equation

u + v = const. gives an ellipse and the equation
u - v = const. gives a hyperbola.

We now form the equation $uv = k$ (4)

and $\frac{u}{v} = F$ (5)

where k and F are positive constants and u and v are positive variables. Let d stand for half the distance between points A and B and to begin with we choose

$k = d^2$ in equation (4).

Our task is now to draw the curve corresponding to the equation $uv = d^2$,

or in another form $\frac{u}{d} = \frac{d}{v}$ (6)

With u=v=d we get a point on the curve which lies halfway between A and B on the line through them. How to obtain other pairs of values for u and v is shown in Figure 3.8.10, which builds upon equation (6).

With the aid of a comp we can as a rule obtain 4 poi on the curve for every pair values u, v, symmetrically loca with respect to A and B (could, of course, have let u and exchange places in equation (4 The curve we get is the so-call lemniscate (Figure 3.8.11).

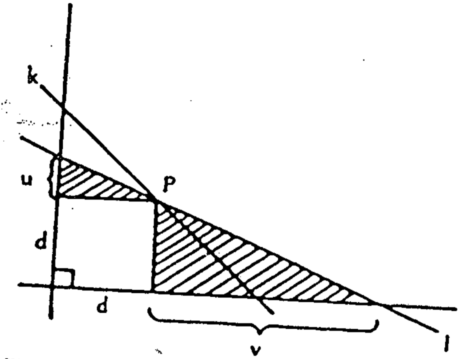


Figure 3.8.10
The shaded triangles are similar. From this follows that $u/d = d/v$ or $uv = d^2$. All lines l through point P satisfy $uv = d^2$. The 45° line gives $u = v = d$.

For k-value less than d^2 the curv is divided up into two ovals. For k values greater tha d^2, we obtain simpl closed curves which are bean-shaped to start with but which straighten out and thereafter turn into curves which remind us of ellipses (Figure 3.8.12).

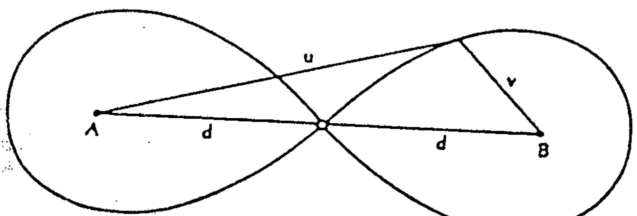


Figure 3.8.11
Lemniscate $uv = d^2$

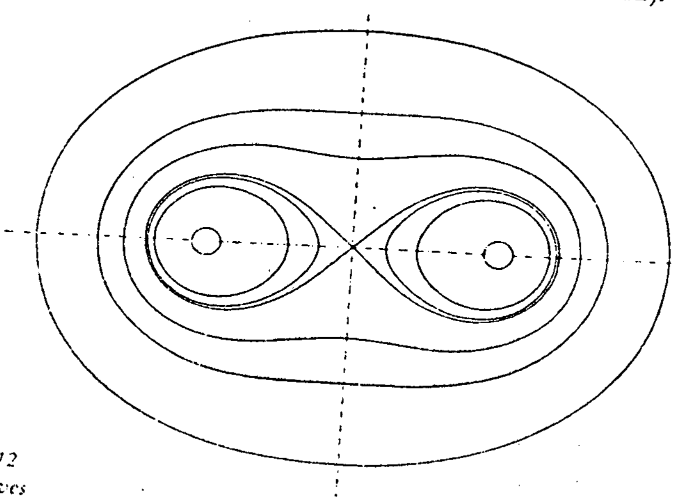


Figure 3.8.12
Cassini curves

The curves obtained are called Cassini curves after the French astronomer J. D. Cassini (1625–1712).

A simple construction, at least when the k-value is not fixed, can be obtained in a manner like that which gave us the ellipse and hyperbola families in Section 3.7.2. Instead of letting the circle radii grow by constant amounts as we did there, we let them increase as a geometric sequence. For example, we might choose the values

$$r = d,\ kd,\ k^2d,\ k^3d,\ \ldots\ (k = \text{a positive constant, e.g. } k = 1.2)$$

$$\text{and} \qquad r = d,\ d/k,\ d/k^2,\ d/k^3,\ \ldots$$

for both u and v values. We then get two sets of geometrically expanding circles about A and B. The A and B circles intersect one another. By connecting points which simultaneously move outward from A and inward toward B we get Cassini curves: the product uv remains constant, since v is divided by k at the same time as u is multiplied by k.

On the other hand, if we connect points which simultaneously move outward from both A and B, we obtain the curves corresponding to equation (5),

$$\frac{u}{v} = \text{a constant F.}$$

How do these curves look? With our construction method here it is not easy to make them perfect but the suspicion definitely arises that it could be a question of circles. A closer analysis shows that equation (5) actually does give circles for $F \neq 1$. They are called division circles, harmonic circles or Apollonios circles (after the Greek who studied them closely). (See further Figure 3.8.13 and Exercise 5.)

When the constant F inequation (5) grows from small values toward infinity, the division circles grow from tiny circles containing A to larger and larger circles, flip over to the other side of the perpendicular bisector of segment AB (when F = 1) and become circles "containing" B[1]. The constant value F = 5 gives the same circle "containing" B as the circle for F = 1 / 5, containing A. Finally, as $F \rightarrow \infty$, the circles enclose B all the more tightly . Figure 3.8.14 shows this family of division circles.

But where is the quality of transformation in this family of circles? Is it not simply a matter of a circle which grows in size and simultane-

[1] It is actually more natural to continue seeing B as an external point to the circle. Refer to the discussion of the "projective" viewpoint which follows later in this section.

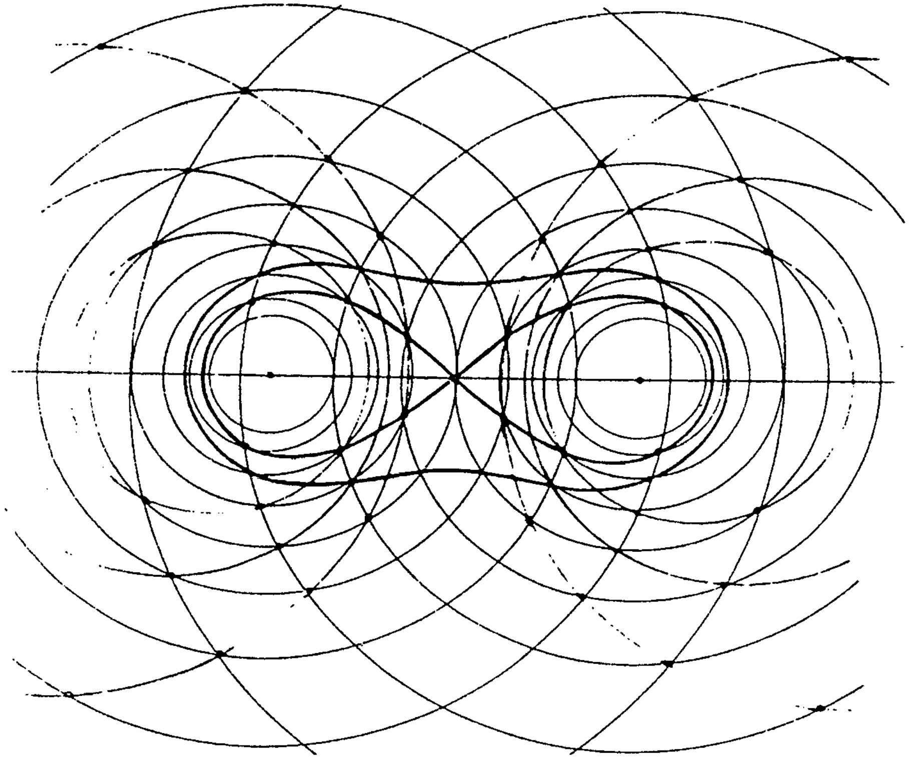


Figure 3.8.13
Two circle families with radii in geometric proportion (k = 1.27). The figure shows two Cassini curves and three pairs of harmonic circles.

ously changes position? The transformation quality first comes to light when we see the family of circles from the perspective of projective geometry. In the next section we will take up some of the basic ideas of *projective* geometry. Here we begin simply by introducing the *projective* line.

In classical Euclidean geometry a line extends infinitely in both directions. On any given axis numbers can be as large as we like, both in the positive direction and in the negative. The line is endless in both of its directions. A projective line arises from the Euclidean when we append an infinitely distant point to the line (and on the line).

Figure 3.8.15 shows a given line a, a given point p above the line and a line p through point P, intersecting line a at point X.

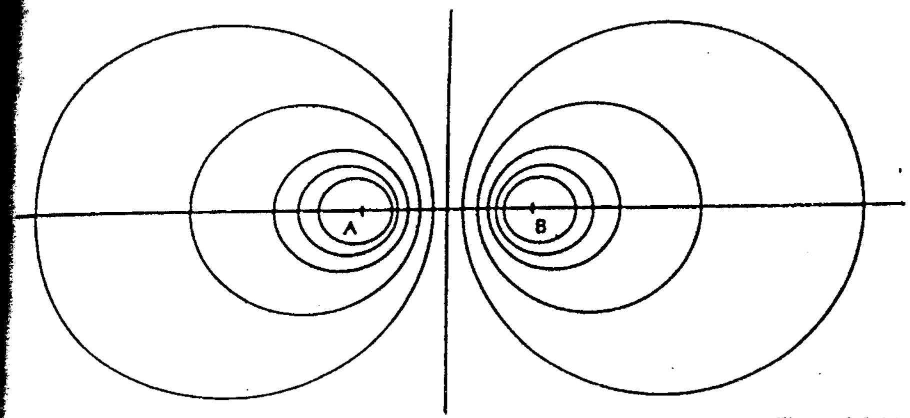


Figure 3.8.14
Harmonic circles corresponding to
F = 1/5, 1/4, 1/3, 1/2 and 2/3 (the largest circle) and to
F = 5, 4, 3, 2 and 3/2 respectively.
F = 1 gives the perpendicular besector to AB.

As line p turns counter-clockwise about P, X moves to the right on line a. When p becomes parallel to a, then according to classic geometry, there exists no point of intersection between them. According to the concepts of projective geometry, we ascribe a common point to lines a and p, namely a "point at infinity," A_∞, on line a. Line p too has a point of infinity, P_∞. Since p runs parallel to a, P_∞ and A_∞ coincide and make up the common point of p and a.

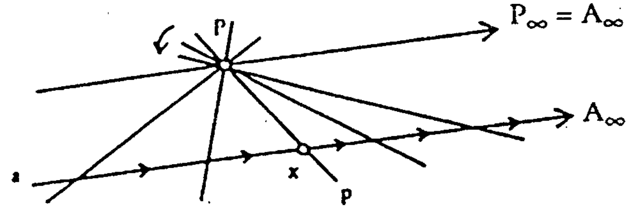


Figure 3.8.15

As P continues to rotate, the intersection X *passes* through A_∞ and then returns from the left at a finite distance. X moves continuously in the same direction, always to the right. That X goes off to the right, through line a's point at infinity and thereby comes back from the left is an idea we cannot clothe in physical form. Considered physically the idea is grotesque. In projective geometry, however, it has a function to serve.

Reconsidering now the growth of harmonic circles from a projective viewpoint, we can describe the circle center's movement with increasing factor F as follows: the center moves to the left on line AB, becomes the point at infinity on AB when F = 1, and returns in from the right as F grows larger than 1.

For F = 1 the circle's interior is the half-plane to the left of the perpendicular bisector to AB. When the circle "flips over," its "interior" — if we wish to maintain continuity — is that which we would normally call the region exterior to the circle. Compared to our habitual way of looking at things, this implies an essential transformation. Let us go a little further with the following example.

Example 4: We let A and B be two fixed points. To begin with, they are the end-points of the diameter of two coincident circles c_1 and c_2. Let n denote the normal bisector to segment AB, and let M_1 and M_2 be the centers of the respective circles (Figure 3.8.16).

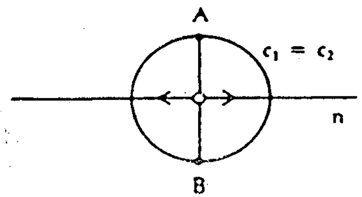


Figure 3.8.16
$M_1 = M_2$ = Center

We now let the centers move at the same rate in opposite directions along line n, M_1 to the left and M_2 to the right. We also require that the circles c_1 and c_2 always pass through the fixed points A and B. This leads to the growth of the circles and to their beginning separation from one another. The plane is hereby divided into various regions.

Let us classify these regions as follows:

No shading:	area not covered by either circle
Horizontal shading:	area covered by c_1 only
Vertical shading:	area covered by c_2 only
Checkered shading:	area covered by both c_1 and c_2

(See Figure 3.8.17).

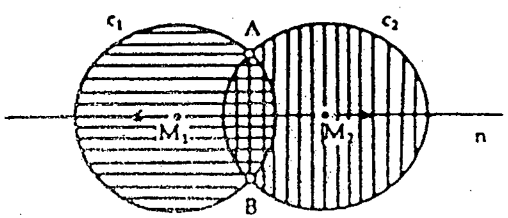


Figure 3.8.17

How does this picture of the plane change, as M_1 and M_2 move away from each other along the projective line n?

Figure 3.8.18 shows the result. In this transformation there arises a quality reminiscent of the negative-positive quality we are familiar with in photography. But it entails considerably more than that, namely, a questioning of the whole concept of inner and outer, shaking up our time-worn habitual thinking that "inside is that which lies within."

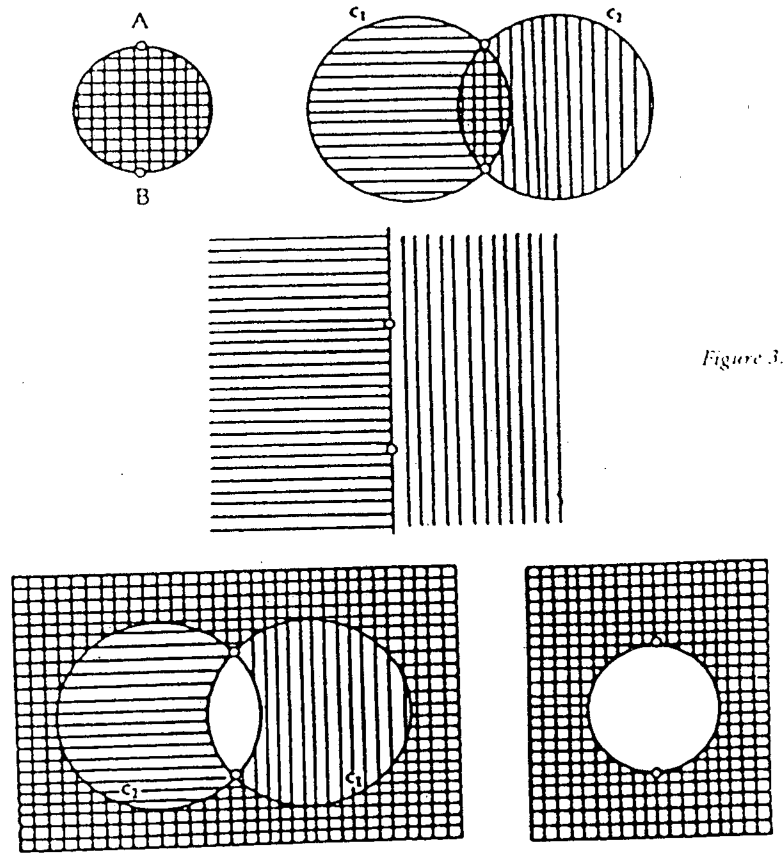


Figure 3.8.18

3.8.2 Forms in Nature, Revisited

Familiarity with geometrical metamorphoses can sometimes shed light on differing natural forms where we would least expect similarity or polarity. This is exemplified by Figure 3.8.19 showing craniums of a

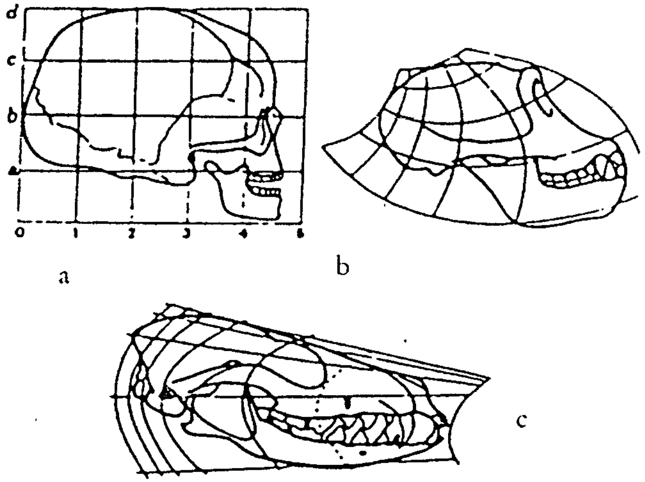


Figure 3.8.19

human, a chimpanzee, and a dog. (The figures are taken from d'Arcy Thompson's classic work "On Growth and Form".)

An interesting polarity may be found comparing man's head and his thigh bones: in the head we find the hard bone on the outside and the soft parts inside the cranium. With the leg it is just the opposite: the soft parts are found outside around the bone. The cross sections of the head and thighbone are very nearly circular and hyperbolic, respectively. A comparison with Example 2 above lets us guess that skull and legbone are each other's opposites in the same way that the interior of a circle is on the inside while interior of a shallow hyperbola, in contrast, is found on what we would normally tend to call the hyperbola's "outside."

3.8.3 *Exercises*

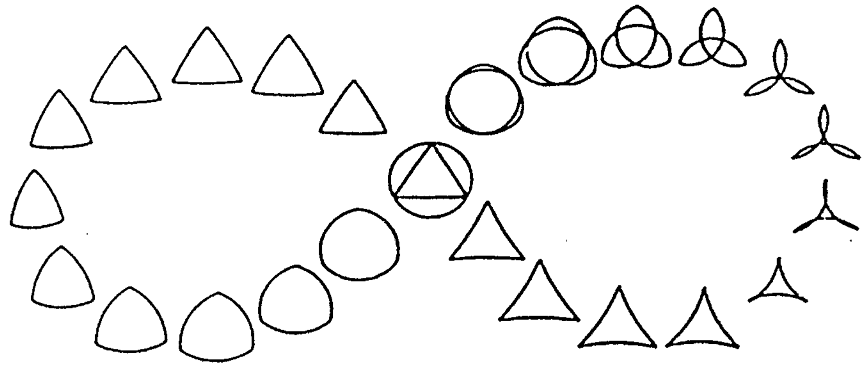
1. We look back at Example 1 in Section 3.8 and choose the distance XY so that Y is on the other side of P, opposite line a. Draw the curve which point Y describes, for different values of segment length XY.

2. What limiting *form* does the conchoid curve of Example 1 approach, as segment XY grows unboundedly?

3. Is there also a limiting form in Figure 3.8.2, as length XY grows unboundedly?

4. Generalize the transformation of Figure 3.8.18 to include three circles. At the start all three circles coincide. If A, B and C are three fixed points on the common circle, located such that they form the corners of an equilateral triangle, then the three circle centers are to move outward along lines passing through the mid-points of the triangle's sides.

The sketch below shows how the "overlapping figure in the middle" (the intersection of the three circles) changes in form.



5. Prove that the division curves of Example 3 in this section truly are circles.

3.9 Spherical Geometry

3.9.1 *A Few Basics*

In their tenth year in the Waldorf schools pupils learn to command the triangle theorems of trigonometry which make it possible to calculate the sides and angles of any triangle we like, when sufficient data is known. Work in geometry is crowned with a period of triangulation out in the field. We have, in a certain sense, complete "command" of plane geometry: we are no longer bound to right triangles and other special cases.

In the eleventh year, during a shorter period, we tackle the sphere — preferably after an introductory study of some important surfaces of revolution, such as the cylinder and the cone. We go here directly to the question: how can one determine distance on a spherical surface (e.g. a ball)? Two points on the sphere are given to us. How do we find the distance between them? Here no pupil has to pay the price of asking, "What use is all this?" — a question which usually is a sign that the teaching has not been concrete enough. The application lies clearly before us: how does one determine the distance between two places on the earth?

Another important problem concerns navigation. On board a ship or airplane, how can one with simple tools determine one's current position?

Other important applications of spherical geometry have to do with the position and movements of the sun, the planets, and the fixed stars, or with the areas of regions on earth, e.g. rounded polar caps or four-sided zones on the earth. Some of the American states have borders which fall along meridians (longitude circles) and latitude lines (circles of constant latitude); the earth's climatic belts also form zones and caps whose areas are of interest.

In short, here are problems and methods worth studying. The most important question during this partial study period, however, is this: how should one build up a geometry for the sphere, starting from basics?

On a plane we measure the distance between two points along the straight line joining them. Such lines serve as "distance lines," so-called geodesic lines. What kind of curves would geodesic lines form on a sphere?

We take a large rubber ball to our aid and mark off two points on its surface with chalk. We turn the ball so that one of the points comes to the top. Now, how does the shortest path to the other point go? "Along a circular arc which you draw straight down — south, if the first point is the North Pole." We draw this arc in by free-hand and discuss whether the answer is correct. The discussion doesn't give a strict proof, but it casts so much light on the problem that the solution with the southerly circle arc seems unique and clear. What radius does the circular arc in question have? The same radius as the sphere. If we lay a circular arc with *smaller* radius through two points, will the distance along this arc be greater? Yes, we see that. (Figure 3.9.1).

The shortest distance between two points on the sphere follows the so-called great circle arc between the points (the great circle has maximum radius = the radius of the sphere). If the points lie diametrically opposite one another, as the Poles, for example, then infinitely many half great circles go between them. They are called meridians when they go from pole to pole. Through Greenwich near London goes the zero-meridian, which everywhere marks off 0° longitude. Together with the equator, which marks 0° latitude, the zero meridian forms the basis for a network which allows us to specify the location of places and the position of ships and airplanes. As shown in Figure 3.9.2a, longitude λ increases eastward to 180°, while westward we let it decrease to -180°. The meridians for 180° and -180° lie precisely opposite the zero meridian.

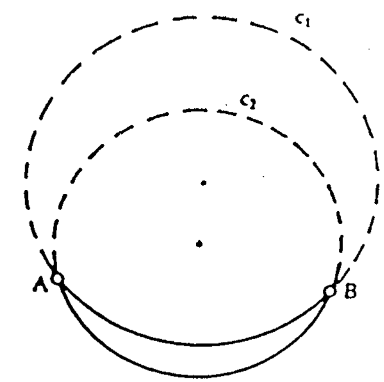


Figure 3.9.1
c_1 = great circle

Latitude φ increases northward to 90° at the North Pole and decreases southward to -90° at the South Pole (Figure 3.9.2b).

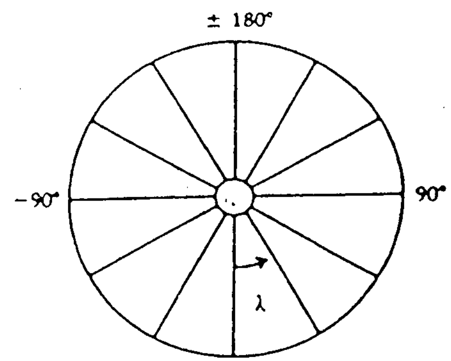


Figure 3.9.2a

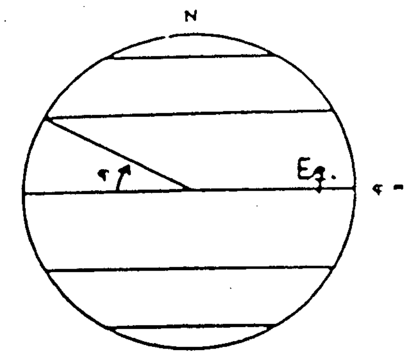


Figure 3.9.2b

With relatively good accuracy we can represent the earth as a sphere whose great circles are 40,000 km long. On the meridians and at the equator, 1 degree corresponds to

$$\frac{40000}{360} \text{ km, or rounded off, 111.1 km.}$$

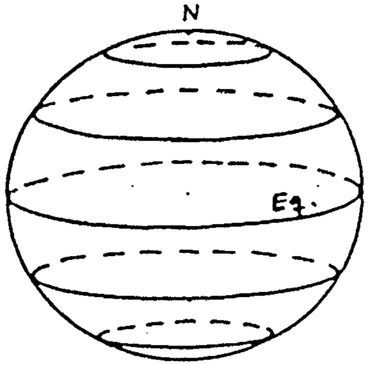


Figure 3.9.3

The nautical mile, a measure of distance at sea, is 1/60 of a degree of length, i.e.

$$\frac{111.1}{60} \text{ km, which gives } 1.852 \text{ km.}$$

Let us now return to the building up of spherical geometry and ask: how do parallel lines run on the sphere? First of all we must observe that all great circles intersect one another. No pair of parallel great circles exists. What curves could then run parallel with a great circle? Clearly a family of circles with smaller radius than the great circle.

Curves parallel to the equator are called naturally enough parallel circles and give constant latitude. They are also called lines of latitude (latitude = width) in this context. (Figure 3.9.3.)

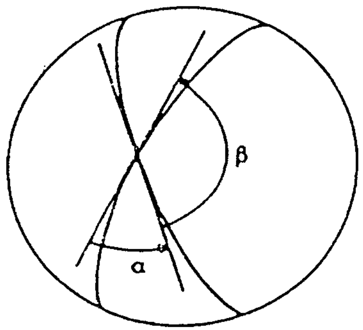


Figure 3.9.4
The angle between two great circles.
$\alpha + \beta = 180°$

When two great circles intersect each other, they do so with a specific angle, which is determined by the angle between the tangents to the two circles at the point of intersection. This is a definition which quite naturally corresponds to the definition of angle between two curves in a plane (Figure 3.9.4.).

Where two great circles intersect, two adjacent angles are formed which together make up 180°.

When 3 great circles intersect one another there arise a number of *spherical triangles.* How many? There are actually 8 formed. Two great circles form four "2-cornered" polygons (Figure 3.9.5-6).

For angles in a spherical triangle, we allow all values under 180°, just as in plane triangles.

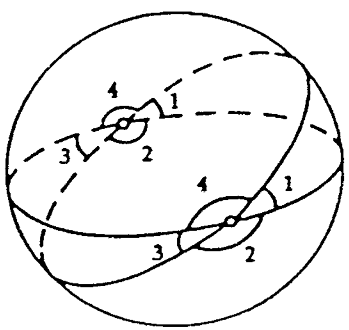
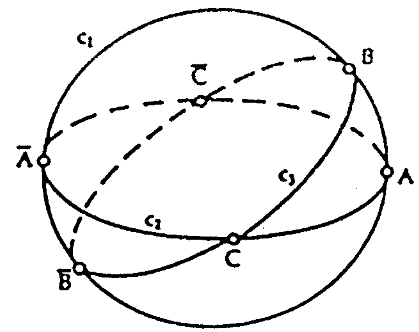


Figure 3.9.5
Two great circles form four "2-cornered" polygons.

Figure 3.9.6
Three great circles c, c and c form 8 triangles:
On the front:
1. ABC
2. $\bar{A}BC$
3. $A\bar{B}C$
4. $\bar{A}\bar{B}C$
On the "back" side:
1. $\bar{A}\bar{B}\bar{C}$
2. $A\bar{B}\bar{C}$
3. $\bar{A}B\bar{C}$
4. $AB\bar{C}$



3.9.2 *What is the Sum of Angles in a Spherical Triangle?*

Isn't it 180°? This is asserted by some pupils, referring to the old familiar statement that "the sum of angles in a triangle is 180°."

But does this theorem apply even on a sphere? Soon an opposition gathers strength and points out the spherical triangle comprising a one-eighth section of the whole sphere and which has right angles in all corners (Figure 3.9.7). Its sum of angles is in fact 3 · 90° = 270°! "Of course, one of the angles can go up to almost 180°. The triangle is then close to a quarter sphere (Figure 3.9.8) and the sum of angles is near to 180° + 2 · 90° = 360°."

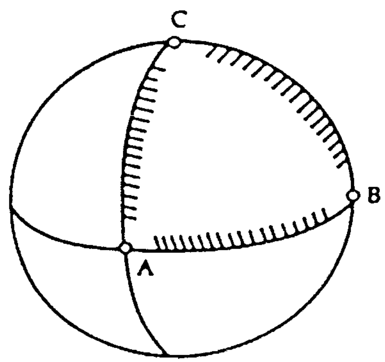


Figure 3.9.7
Sum of angles 270°

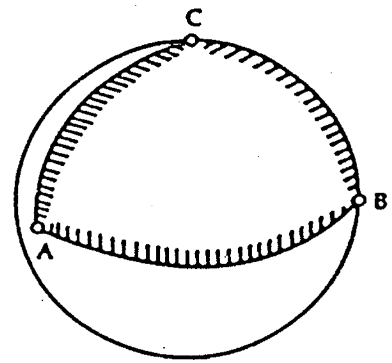


Figure 3.9.8
Sum of angles near 360°

Can the sum be even larger? (The teacher is never satisfied!) Fantasy now begins to wane, but someone suggests that our "almost-1/4-sphere" ought to be able to be expanded downward and include even more area. What happens if we extend the two short sides of our near-360°-triangle downward and let the long bottom edge instead shrink smaller (Figure 3.9.9)?

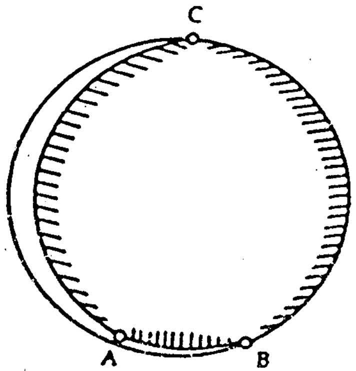


Figure 3.9.9
Spherical triangle with sum of angles near 3.180° = 540°.
It covers almost half the globe.

Will not the two 90° angles then become larger?

We make a paper pattern as in Figure 3.9.10. It is simply a circular disc where we have cut away an insignificant "pie" wedge, whose center angle is small, say 3°.

We divide this big circular disc into three equal parts (each with 89° angle), draw radii and fold up the paper into a shallow cone (with point downward). The three arcs then form the sides of a spherical triangle, which in our imagination is a hemisphere arching high over the cone corner (Figure 3.9.11). As can be seen, the angles in the spherical triangle we get will all be nearly 180°.

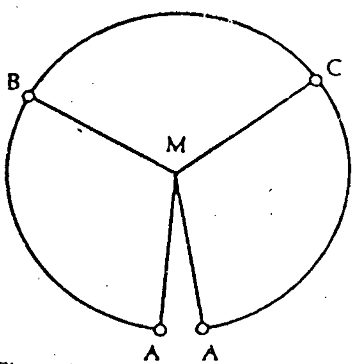


Figure 3.9.10

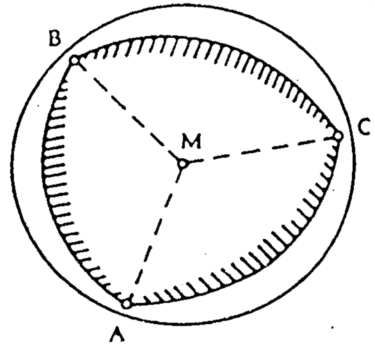


Figure 3.9.11

The conclusion is that the sum of angles in a spherical triangle can vary from slightly more than 180° (a small triangle, flat and almost in the plane) to slightly less than 540° (a triangle covering almost half the sphere).

Closer investigation of the sum of angles (V degrees), which we do not go into here, reveals a linear relationship with the area of the triangle (A):

$$A = \frac{\pi R^2}{180}(V - 180) \qquad \text{where } R = \text{the radius of the sphere} \qquad (1)$$

and turned around

$$V = \frac{180\, A}{pR^2} + 180 \qquad (2)$$

Formula (2) can be written

$$V - 180 = \text{constant} \cdot A$$

or

$$E = \text{constant} \cdot A,$$

where E = excess above 180° of the sum of angles, and the constant is

$$\frac{180}{\pi R^2}$$

The excess angle E is thus proportional to the triangle area. (See diagram 3.9.12).

On a sphere, therefore, a triangle's sum of angles is uniquely determined by the area or vice-versa, the area is determined by the sum of the angles. In the plane, the sum of angles is fixed at 180°, while the area can vary freely.

Do similar triangles exist on the sphere (other than congruent triangles)? In other words, for a given triangle, does there exist a smaller or a larger with the same angles? The answer which we readily see is "no." The greater the area is made, the greater is the sum of angles.

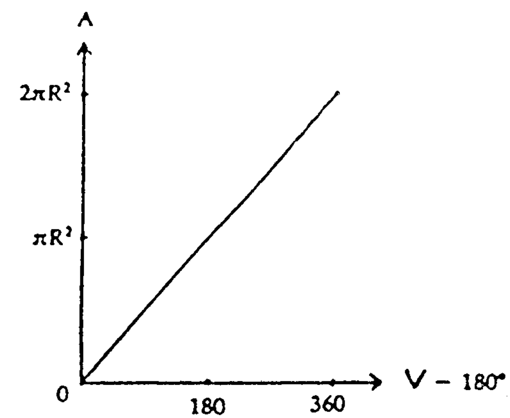
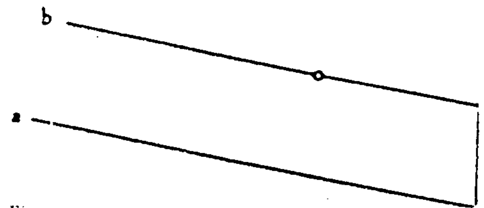


Figure 3.9.12

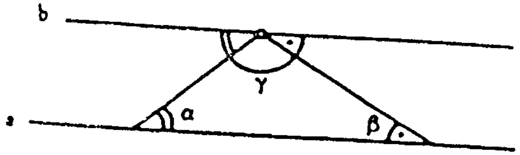
$A = \frac{\pi R^2}{180}(V - 180)$
Area is proportional to the "excess" angle sum V - 180°

That the sum of angles in a plane is always 180° is closely related to the axiom on parallel lines, i.e. with the existence of a unique parallel (b) to a given line (a) and passing through a specific point off of the line a (Figure 3.9.13).



We do not have this axiom for the sphere, and thus the conditions are completely different in regard to the sum of angles in a triangle.

Figure 3.9.13
$\alpha + \beta + \gamma = 180°$ in the plane.



3.9.3 *Distance*

We shall now take up a practical problem: determining the distance between two points on the earth.

With the aid of spherical trigonometry the calculation is quite easy; we would simply use the longitudes and latitudes of the points and plug their values into a formula. But it would take time to develop trigonometry up to the point of the formula which is needed .... We will here use a method with compass, protractor and straight-edge which provides a beautiful example of how one with the aid of methods taken from plane geometry can master the surface of the sphere. As we will see, the method depends upon the simple fact that the circle is a plane curve and that every plane cutting the sphere does so along a circle.

Let us determine the distance between Stockholm-Arlanda Airport (long. 18°; lat. 60°) and Buenos Aires (long. -58°; lat. -34°).

To begin with we note that the longitudinal difference between Stockholm (S) and Buenos Aires (B) is:

$$18° - (-58°) = 76°;$$

S lies thus 76° east of B (not, of course, straight east).

We draw a circle C according to Figure 3.9.14 and let it represent the great circle through B, the western most of the two cities.

The horizontal diameter of the circle is a projection of the equator, and at the top we have the North Pole, N.

We now imagine replacing half of the globe (the half with Stockholm on it) back over the circle so that the North Pole coincides with N and Buenos Aires with B. Stockholm finds itself somewhere in the air above the paper.

We shall project Stockholm straight down onto the paper and mark the point with an S. How do we do this? First we draw the horizontal chord corresponding to Stockholm's circle of latitude, i.e. we project the 60° circle of latitude down to the paper — line k in Figure 3.9.15.

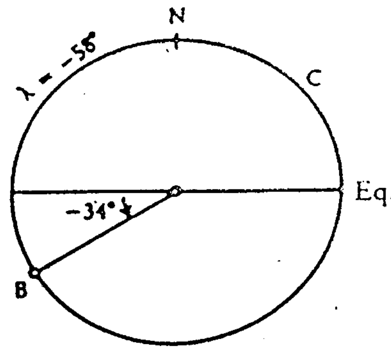


Figure 3.9.14
B = Buenos Aires, N = The North Pole

With k as an axis we rotate half of the latitude circle into the plane of the paper (above the North Pole) and obtain the half circle c. On c we move 76° east from B's meridian (the left half of the great circle) by placing the protractor at the mid-point of chord k and marking off 76° to the east.

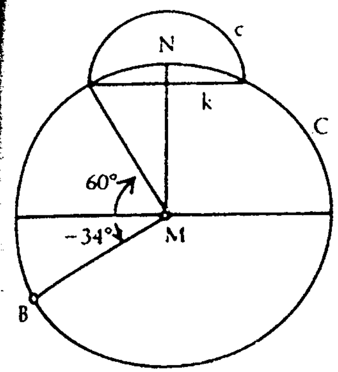


Figure 3.9.15

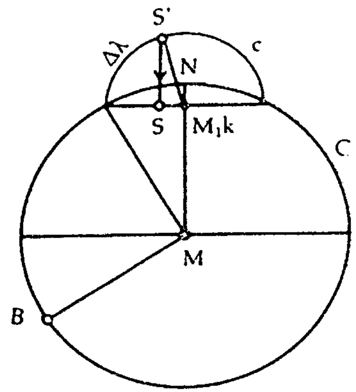


Figure 3.9.16
$\Delta\lambda = \lambda_S - \lambda_B = 76°$

We then come to the point S´ on c, which is Stockholm's location on the paper after we have flipped the latitude circle down onto the plane of the paper. If we now rotate the latitude circle back up to its position on the globe, how will S′ move projected onto the paper? At right angles to k.

We therefore draw a line at right angles to k and obtain Stockholm's position (projected onto paper). (Figure 3.9.16)

We next imagine the shortest route from Stockholm to Buenos Aires. In reality it is a great circle arc. What we want to know is: how many degrees of arc, how large an angle, does this great circle take up at the center of the globe, at M? On paper its projection would be an ellipse arc, but could we possibly rotate the globe so that the route comes to lie in the plane of the paper?

Yes, it can be done. We rotate the globe about the diameter BM, i.e. we use BM as an axis! Stockholm then moves down toward the paper, and in projection, along that line which goes through S at right angles to BM, coming to the new position S″ on the great circle C (Figure 3.9.17).

Finally we measure the angle

$$v = BMS''$$

and obtain the *angular distance* for the shortest route BS. The protractor shows 113°.

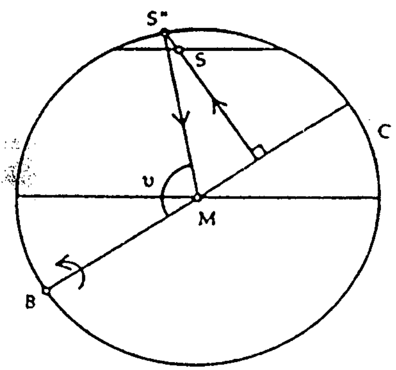


Figure 3.9.17

Since each degree on a great circle corresponds to 111.1 km, the distance is 113 · 111.1 km, or approximately 12,600 km.

If both cities have the same longitude, the construction is unnecessary. The one city then lies straight south of the other, i.e. the cities lie on the same meridian and we can directly calculate the distance. For example, the distance between Stockholm and Cape Town (18°; -34°) is quite simply the latitude difference times 111.1 km, i.e.,

$$|60 - (-34)| \cdot 111.1 \text{ km} \approx 10{,}400 \text{ km}.$$

In class, exercises on this theme are somewhat embellished with flight times, local times, landing times, etc. The class has available a world atlas showing local times, and the pupils may easily be left to themselves finding the longitudes and latitudes of the cities in question.

3.9.4 Exercises

1. How far is it from Stockholm Arlanda (18°; 60°) to the poles? (A great circle is 40,000 km.)

2. How far is it *straight east* from Cape Town (18°; -34°) to Sydney (151°; -34°)?

3. Determine with compass and protractor the distance between
   a) Stockholm (18°; 59°) and Mexico City (-99°; 19°)
   b) Philadelphia (-75°; 40°) and Dar es Salaam (39°; 7°)

4. Calculate the distance from New York (-74°; 41°) to Hanoi (106°; 21°).

5. A person who is *not* at the North Pole travels first 1000 km south, then 1000 km straight east and finally 1000 km north. He is then back at his starting point. Where on earth is he?

6. How large must the sum of angles in a spherical triangle be for the triangle's area to cover 25% of the sphere?

7. The formula for calculation of the area of a spherical cap or a ring zone on a sphere is

$$A = 2\pi Rh,$$

where R is the radius of the sphere and h is the width of the ring zone or the cap.

What per cent of the earth's area is taken up by the temperate zones, whose latitudes vary between 30° and 60° (both above and below the equator)?

3.10 A Little About Projective Geometry

3.10.1 A Projection Problem

In Euclidean geometry, the lengths of line segments and the size of angles play a decisive role. To become aware of this we have only to remember a few basic problems and theorems from Euclid's work, the *Elements:*

- Constructing a given angle with compass and straight edge, i.e. reproducing an angle
- Constructing a chord of given length through a specified point within a circle
- Sides opposite equal angles in a triangle are equal; opposite a larger angle is a larger side
- The Pythagorean Theorem: in a right triangle, the square of the hypotenuse is equal to the sum of the squares of the sides which form the right angle (Figure 3.10.1).

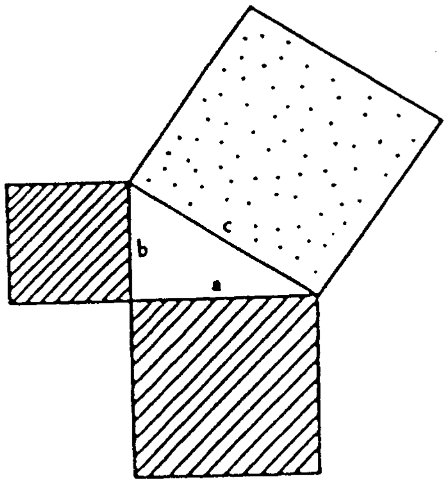


Figure 3.10.1
The Pythagorean theoriem: dotted area = sum of the shaded areas. $c^2 = a^2 + b^2$

There are, however, many problems in geometry which do not concern quantities but rather put the stress on relations. We know that if an object casts a shadow, then the shadow, the object and the light source stand in certain relation to one another. The shadow picture is dependent on the contours of the object, on the location of the light source, and on the surface upon which the shadow is cast.

Since shadow plays a role in drawing and painting, there eventually developed a "theory" of shadows. Earlier still the need for drawing objects in perspective had arisen among artists. The first attempts at this in the 1400's sometimes appeared rather comical: one can see a table with plates drawn such that they seem to be sliding down the table toward the viewer. People sitting in chairs appear to be half-sitting, half-standing; the seats of the chairs slope steeply forward in the picture.

Many other examples could be presented. Stronger and stronger was the desire to be able to draw pictures in perspective so that the objects appeared natural. And from the theory of perspective developed

projective geometry. In fact, it was projective geometry which laid the foundation for the theory of shadows. The name itself, projective, comes from the Latin word *projicere*, to throw or cast, verbs which we in fact use concerning shadows: the sun throws (casts) a shadow.

The simplest projections prevail between two plane figures, e.g. two triangles. Let us pose the following problems:

1. Can a point-source of light and a small equilateral triangle be positioned so that the triangle's shadow precisely covers a given larger triangle with a specified position in the room?

2. Can one find a plane and a vantage point for the eye in the room, where the perspective picture of a given triangle on the plane will be a specified small equilateral triangle?

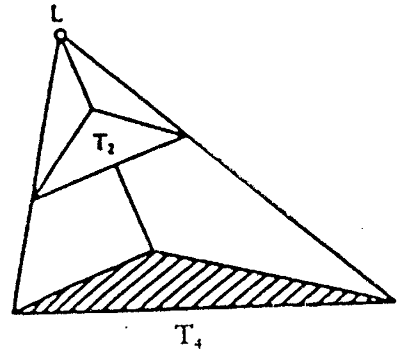


Figure 3.10.2
T_1 = given triangle
T_2 = equilateral triangle

Figure 3.10.2 illustrates these problems (with the desired conditions fulfilled). What we see directly in this figure is that the two problems are the same from a geometrical point of view. They can be formulated in the following way: A triangle T_1 is given in 3-dimensional space. Further, a small equilateral triangle is given. Can this latter triangle and a point be so positioned that both triangles become perspective in relation to the point, regardless of T_1's shape?

To begin with, we put a large paper triangle (of arbitrary shape) on the classroom floor and cut out small equilateral triangles of cardboard. Some pupils instead cut equilateral triangular holes in the cardboard. We now experiment with our eyes to see if we can place the little triangle (cut-out or hole) in front of the eye in such a position that the cut-out "exactly" covers the triangle on the floor or that the large triangle could "just barely" be seen through the hole. We have to move around a bit but finally seem able to find positions in which we succeed. No one doubts that the problem has a solution. On the other hand, our method could certainly not be called a proof.

How could we achieve a proof for the positive answer?

We begin to simplify the problem by choosing a special case. Why not first let the given large triangle on the floor be equilateral, i.e. have the same form as the little triangle?

Now it is easy to devise a construction which proves the answer: the small triangle can be placed "horizontally" directly above the triangle on the floor so that corresponding corners of the two triangles can be connected to form a three-sided pyramid. We draw a picture of this reasoning and conclude not only that the idea holds water but also that the little triangle can be as small as we like. The point source of light (the center of perspective) comes to lie at the pyramid's top (Figure 3.10.3).

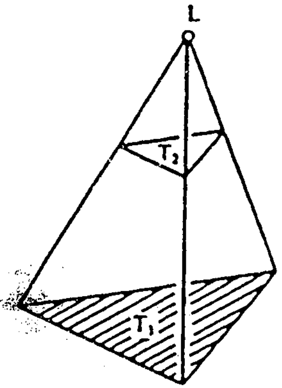


Figure 3.10.3
A simple case: T_1 and T_2 are equilateral.

We return now to the original problem and try to find other simplifications. Since the size of the small triangle does not matter, could we exchange it for a larger equilateral triangle whose sides are just as long as one of the sides of the triangle on the floor?

"Then it'll be easy! We can put the equilateral triangle right down on the floor with one of its edges alongside an edge of the floor triangle."

We implement this thought-process with paper triangles and draw a sketch. But where shall the lamp be put? It takes a good while in fact before the class figures this out. They are apparently locked in by the thought that there is only *one* place where the light can be put and are surprised that it can be placed anywhere at all along the "upper" part of line a in Figure 3.10.4.

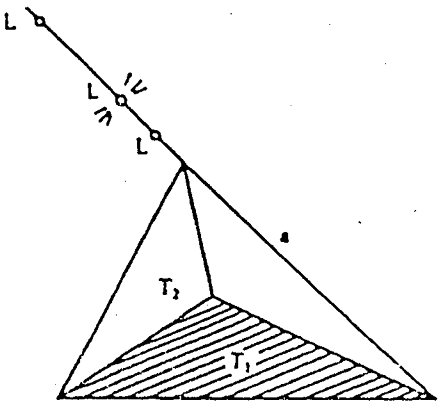


Figure 3.10.4
Sketch of the solution for a simplified case.

And once again, we return for a fresh attack on the original problem. Now there is use for the knowledge gained earlier that the small triangle is unimportant in the first simplified case.

Might not the equilateral triangle of Figure 3.10.4 be traded for a smaller? "Naturally, it can be replaced by a smaller one nearer the light."

"How, in that case?"

"We can shift the triangle, parallel to itself, toward the light. It will then get smaller but keep its equilateral shape."

The problem is solved and we draw the final figure (Figure 3.10.5).

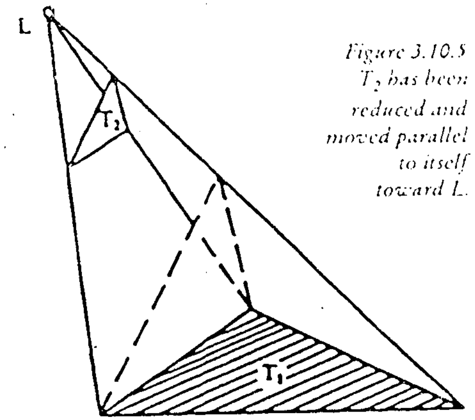


Figure 3.10.5
T_2 has been reduced and moved parallel to itself toward L.

3.10.2 *Désargues' Triangles*

Gérard Désargues (1593-1662), a French architect who was active in Paris and Lyon, published a theorem on perspective triangles in the year 1648. This theorem has come to be one of the foundations of geometry and plays a decisive role particularly in the theory of perspective. Désargues wrote a dissertation on perspective. His work was appreciated by none less than Pascal, who investigated closely related problems. Désargues' and Pascal's names have come to be associated with respectively the theorem on perspective triangles and an important proposition on hexagrams (6-sided polygons), inscribed in conic sections. Both of these mathematicians laid important foundation stones in projective geometry but their contributions were not given any particular attention by their contemporaries. No trace was left of Désargues' old publications, but in the nineteenth century an English mathematician, Arthur Cayley, succeeded in finding an old transcript of Désargues' manuscript. Not until then were Désargues' and Pascal's methods properly appreciated.

We shall not go into the structure of projective geometry further than to acquaint ourselves with Désargues' triangles and with a few simple so-called dualities. We can begin directly from our introductory study with the paper triangles.

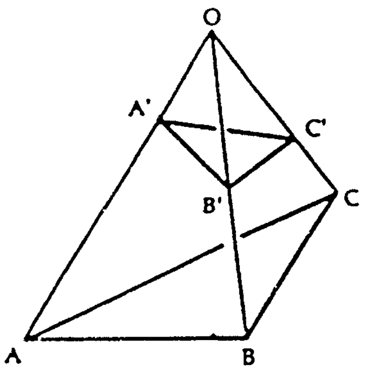


Figure 3.10.6
Two triangles perspective with respect to the point O.

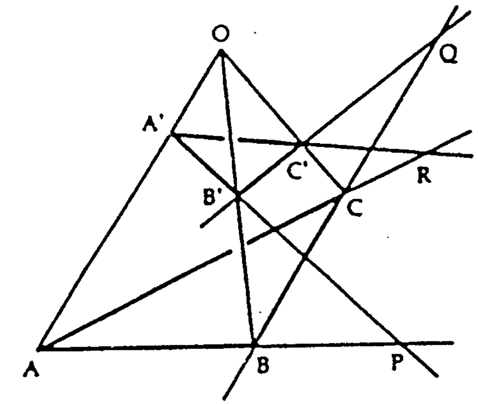


Figure 3.10.7

We draw two triangles in perspective positions (Figure 3.10.6). ABC is perspective with A´B´C´ (and vice versa)when lines AA´, BB´ and CC´ go through a common point, the center of perspective (O).

Problem: Have these triangles any other lines in common than the three lines through O? Obviously not, if we keep to the figure we have drawn. But let us consider the triangles as each being formed by three *lines* and ask if these lines meet somewhere in space (Figure 3.10.7).

Some pupils answer "yes" immediately, for one pupil perhaps based on a feeling, but others can motivate their answers: lines AB and A´B´ must (in general) intersect each other, since they lie in a common plane, the "wall" OAB of a pyramid. For the same reason must BC and B´C´ intersect, as well as lines AC and A´C´ (in general). We have thus obtained three points of intersection in space, which we will call:

$$P = AB \times A'B'$$
$$Q = BC \times B'C'$$
$$R = CA \times C'A'$$

The class extends the lines and marks the intersections; for some pupils one or more of the intersections turn out to be off the paper. With time one becomes better at choosing the lines so that everything stays on the paper.

How do P, Q and R lie? "In a straight line," someone answers. "Almost a straight line," say others. "Not mine." "Well, maybe they do." Et cetera.

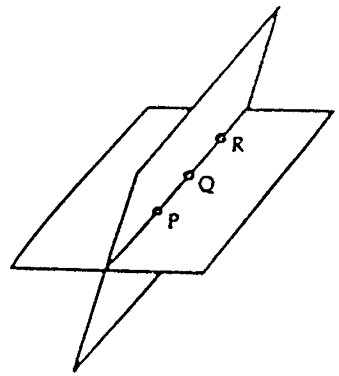


Figure 3.10.8

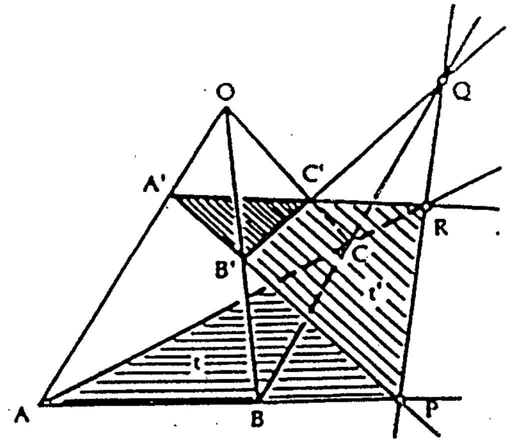


Figure 3.10.9
PQR = the intersection of the planes of the triangles.

We compare each other's figures and find that in those cases where P, Q and R do not lie in a line, they at least form a very long thin triangle.

Can it be that P, Q and R always lie on a line?

We direct out attention to the *planes* t and t´ in which the two triangles lie. The planes t and t´ meet at the three points P, Q and R. How do two planes meet, if they have a point in common? Always along a straight line! (Figure 3.10.8)

We have herewith reached the conclusion that P, Q and R lie on a line and one that particular line which makes up the intersection of the planes t and t´. All that is left to do is to draw a clear figure illustrating this, for example, as in Figure 3.10.9. "But what about if AB is parallel to A´B´?"

Let us draw this case! (Figure 3.10.10a). What happens now with P, Q and R? We can still, in general, construct Q and R. The line QR, or the so-called Désargues' line, will still be determined by these two points. But how does it lie?

Look very carefully! — "It goes parallel to AB and A´B´" (Figure 3.10.10b).

It is not difficult for us to focus on the triangle planes t and t´ and *see* that QR must be parallel to AB. And P? "It doesn't exist," some say. "We could say that P is infinitely far away," say others.

It is precisely the latter way of looking at things which Désargues, and later all work in projective geometry, builds upon.

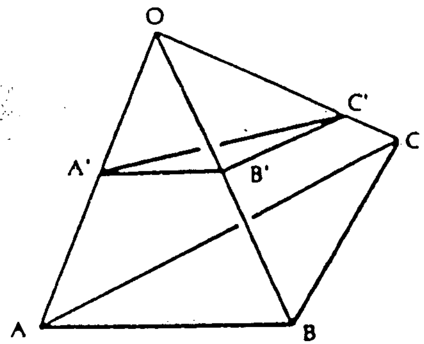


Figure 3.10.10a
AB ∥ A'B'

The point is that it is our free option to give every line one point at infinity, but not two. Were that the case we would violate the axiom that two lines have *one* point of intersection. We would of course be removing the unsatisfying exception that two lines *sometimes* lack a common point (intersection) but land instead in a new exception: parallel lines would then have *two* common points at infinity.

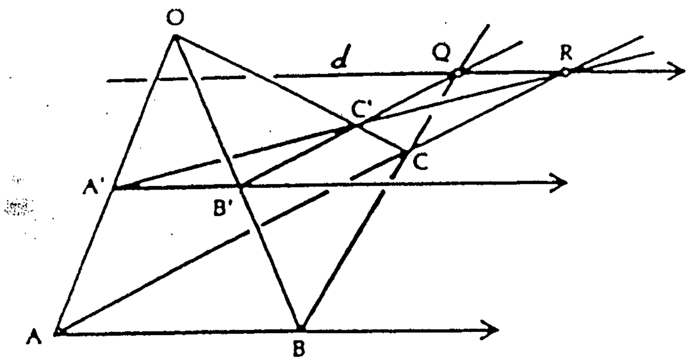


Figure 3.10.10b
d is Désargues' line

The classical Euclidean line has infinite extension in two directions. The new concept which projective geometry introduced was that every line has *one* point at infinity in addition to the finitely distant Euclidean points. One then speaks of the projective line. With the addition of the point at infinity, a line closes on itself: any point X which moves in the one direction of movement along the projective line "returns" from the other side with the same direction of movement at a finite distance, if it passes through the point at infinity, P_∞. Consider, for example, the point of intersection between a line and a rotating line as shown in Figure 3.10.11. This continuity aspect was introduced by the Frenchman Victor Poncelet (1788-1867).

But back to Désargues' triangles. If *two* sides of the one triangle are parallel with two sides (respectively) of the other, what happens to Désargues' line? As one easily sees, all three sides in T are then parallel with the corresponding sides in T'. The triangles then lie in parallel planes and all of the points P, Q and R will be points of infinity. (Figure 3.10.12)

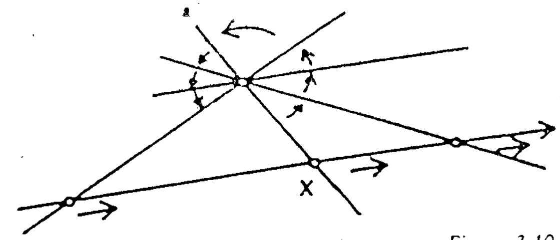


Figure 3.10.11

Can one here speak of Désargues' line? Yes! Analogous to two parallel lines having a common point at infinity, we give parallel planes a common line at infinity.

The question now is to see whether points at infinity and lines at infinity can be fitted in with the basic axioms concerning points and lines which form the foundation of classical geometry and which we do not wish to throw overboard. Here we must skip over most of such a study for reasons of space. However, it turns out that

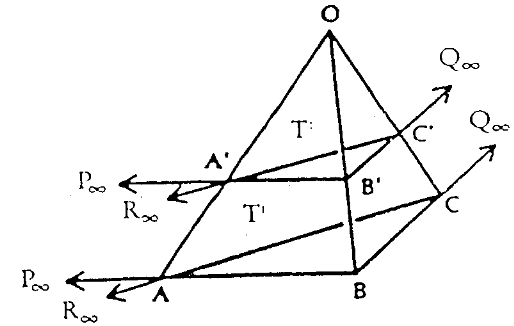


Figure 3.10.12
AB ∥ A'B', BC ∥ B'C'.

- two points have *one* line connecting them;
- two planes have *one* line of intersection.

We observe that two lines (in space) in general lack a common point. But *if* the lines have a common point, then they also have a common plane, and vice versa.

These basic phenomena let us guess that points and planes form a duality in *space*. In the plane, the dual elements are points and lines:

- two points have one line connecting them (common line)
- two lines have one point of intersection (common point)

A particularly important result of this is that plane geometry may be built up with complete symmetry: for every configuration of points and lines, there is a corresponding opposite dual configuration where the role of the points is taken over by the lines and the role of the lines by the points.

Here follow a few simple examples which exhibit such duality:

Three points A, B and C form corners of a triangle, if they do not lie in a straight line (if they lack a common line). Figure 3.10.13a.

Three lines a, b and c form a three-sided figure, if they do not pass through a single point (if they lack a common point). Figure 3.10.13b.

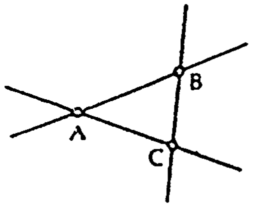


Figure 3.10.13a

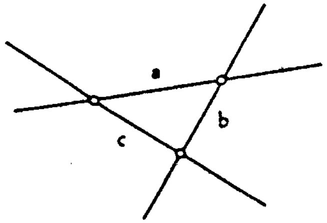


Figure 3.10.13b

A complete quadrangle (four-angled polygon) is formed by 4 points (of which no 3 form a straight line) and the 6 connecting lines determined by the points (Figure 3.10.14a).

A complete quadrilateral (four-sided polygon) is formed by 4 lines (of which no 3 pass through a point) and the 6 points of intersection determined by the lines (Figure 3.10.14b).

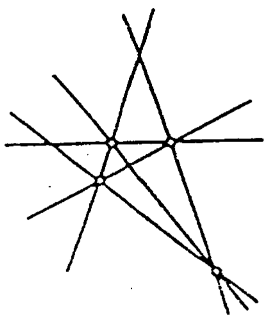


Figure 3.10.14a
Complete quadrangle

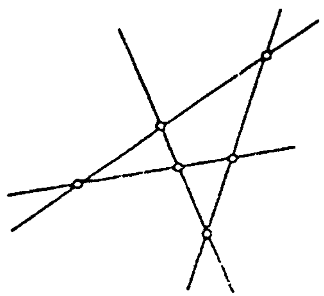


Figure 3.10.14b
Complete quadrilateral

A circle may be considered as made up of infinitely many points (Figure 3.10.15a).

A circle may be considered to be formed by infinitely many lines (Figure 3.10.15b).

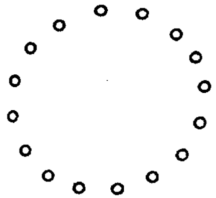


Figure 3.10.15a
A circle can be formed of points.

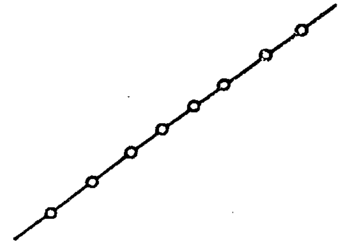
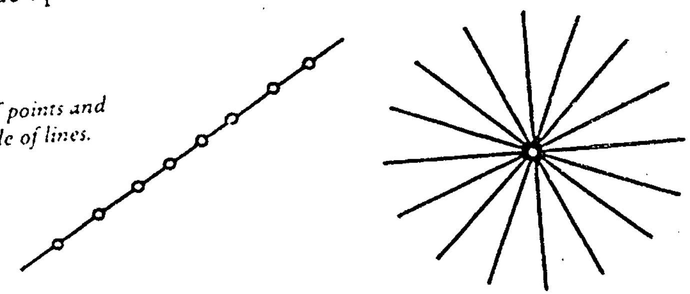


Figure 3.10.15b
A circle can be formed by lines.

The fundamental duality in the plane comes out even with the two basic elements themselves, the point and the line. Not only the point, but also the line may be considered as the basic building block, an indivisible fundamental element. But *if* we consider the line as made up of infinitely many points (a row of points), then we may by the same token consider the point as made up of infinitely many lines, see Figure 3.10.16.

Figure 3.10.16
The line as a row of points and the point as a bundle of lines.



The functional duality between the row of points and the line bundle shows clearly that

— a point-line is created when a line bundle is intersected by a line not in the bundle.

—a line bundle is created when the points of a point-line are connected with a point not on the line.

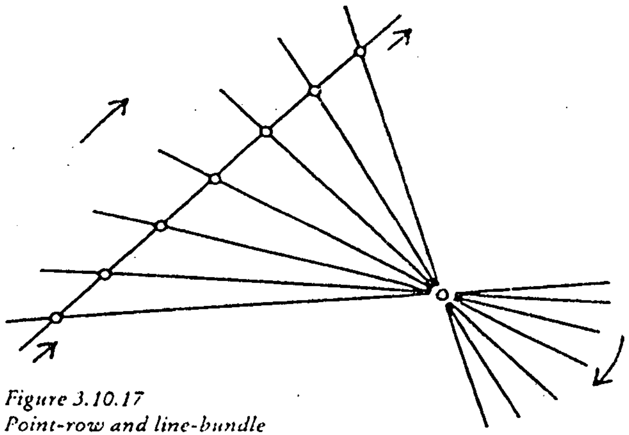


Figure 3.10.17
Point-row and line-bundle

In *space*, as mentioned earlier, the plane and point are dual elements of each other.

Projective geometry's dual structure was not discovered until the 1820's, but it then became the most important factor in further developments which gave projective geometry a dominant position over earlier geometries.

Besides perspective and the theory of shadows, applied projective geometry plays a role in photogrammetry (with applications in the technology for producing maps via photographs, for example). Its great value in the school is that it gives pupils entirely new aspects of what geometry is all about, it allows them to experience considerably more of geometric qualities than the usual quantitative problems, and it for work and discovery to youngsters who do not have, or believe themselves not to have, ability for the equation-solving of analytical geometry.

Particularly the duality exercises are aimed at developing agility in thinking and to mould will power into thinking. One learns to see problems from opposing directions and gains a much broader concept of space than the usual idea of space as a great empty container of points.

3.10.3 Exercises

1. We start with the Désargues figure in Figure 3.10.9 and investigate whether point O, the center of perspective, may be placed at A: does the figure have two triangles which are perspective in space with respect to A?

The answer is actually yes! Find the triangle and determine their Désargues-line.

2. The same problem as Exercise 1 above but with C as the new center of perspective.

3. (a) We draw Figure 3.10.9 once again and investigate whether the figure contains two perspective triangles with the line OB´B as the Désargues-line. Where do the triangles lie and where is their center of perspective?

(b) Show that any one of the 10 points in Figure 3.10.9 can be the center of perspective and that any one of the 10 lines cane be the Désargues-line.

4. In the complete quadrilateral of Figure 3.10.14b we denote the quadrangle's corners by A, B, C and D and the other points by E and F (E = AB x CD, F = BC x DA).

If the pairs of points (A,C), (B,D) and (E,F) are connected we get three lines which form a triangle. What triangle would correspond to this in Figure 3.10.14a?

5. We label four of the points of the dot-circle in Figure 3.10.15a as A, B, C and D (for example, clockwise). Corresponding to this in Figure 3.10.15b will then be four tangents a, b, c and d in cyclical sequence. These four elements determine a complete quadrangle and a complete quadrilateral respectively. Study how these correspond to each other in the two figures. Among other things, compare the positions of the points and the lines in the triangles which, according to Exercise 4, may be associated with the figures.

6. Draw a Désargues-figure with the center of perspective infinitely far off.

7. Draw a Désargues-figure where one of the corners of the triangles is a point at infinity.

3.11 George Boole and Set Theory

3.11.1 A Pioneering Contribution

It is said that Gottfried Wilhelm von Leibniz (1646-1716), one of mankind's universal geniuses, read Latin fluently and began to study

Greek before reaching the age of 12. As a 20-year-old he published the work "De Arte Combinatoria" (On the Art of Combination) which he later considered "a schoolboy's essay" but which came to be the starting point for a new level of abstraction in pure mathematics.

If we hold ourselves to arithmetic and algebra, we can separate out three levels of abstraction, as far as numbers are concerned:

1. The numbers are drawn or written as pictures, for example the hieroglyphs of ancient Egypt. The number 9 is indicated by nine pictures of the symbol which stands for 1, and so on. The degree of concreteness is very high here.

2. Numbers are indicated by quite abstract characters, for example, as in the base-ten system. The symbol 1 for one is concrete, but already the symbol 2 for two is quite different from the ancient Egyptian or Roman II.

The number ten is indicated by a combination of two digits, 1 and 0, and so forth. In position systems, among them our 10-system, each added digit gives different sized contributions depending on its position (place) in the number. For example, the first five in 5157 contributes 5000 to the number, the second five contributes 50.

3. Numbers are represented by letters. For example, 2n means any even number if n is a whole number (an integer). 2n + 1 means any odd number, a variable x can take on any real value — and so on.

Thanks to representation of numbers by letters it is possible to carry out proofs in arithmetic as in Section 3.5.

Leibniz lifted arithmetic up to a considerably higher level of abstraction by letting letters represent *intervals* on the number line (the x-axis of real numbers). He developed, as we shall soon see, the basis for a kind of interval arithmetic in his work "The Art of Combination." His purpose in this was not small: he wished to create "a general method in which all truths of the reason would be reduced to a kind of calculation. At the same time this would be a sort of universal language or script, but infinitely different from all those projected hitherto, for the symbols and even the words in it would direct the reason..."[1]

The purpose and the wording remind one of Descartes' declaration of program in a work on geometry and methods from 1637, where

[1]From E. T. Bell, *Men of Mathematics*

Descartes saw himself as laying out "a completely new science, which will come to admit a general solution of all such problems as can be put as questions of quantities, continuous or discontinuous, each in accordance with its own nature... so that almost nothing will be left to discover in geometry." (Descartes, "Dissertation on the Method of Geometry.")

Following Section XX in one of the fragments of Leibniz' essay, which is still in existence, we let A, B C etc. denote intervals on a line and the symbol ⊕ denote an operation which we can call "addition of intervals." A ⊕ B will then denote the interval which contains points from the interval A or from B or both. We will say, henceforth, "points from A or B" (Figure 3.11.1).

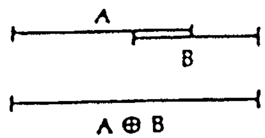


Figure 3.11.1a

Figure 3.11.1b
A ⊕ B includes here both intervals.

Leibniz notes such basic rules for interval arithmetic as

"Axiom 1"	$B \oplus N = N\ \ B$	(as we see, the commutative law for the newly introduced addition)
"Axiom 2"	$A \oplus A = A$	(quite "revolutionary" compared with $a + a = 2a$!)

"Proposition 5" If A is contained in B and A = C, then C is contained in B.
"Proposition 7" A is contained in A.
"Proposition 9" If $A = B$, then $A \oplus C = B \oplus C$
"Proposition 10" If $A = L$ and $B = M$, then $A \oplus B = L \oplus M$.

Apart from "Axiom 2" we recognize these rules from our own arithmetic and algebra. But let us ask, how can we or should we solve an interval equation of the type

$$A \oplus X = A \oplus B \quad \text{where X is unknown?}$$

Can we simply apply our usual equation-solving procedures and write $X = B$?

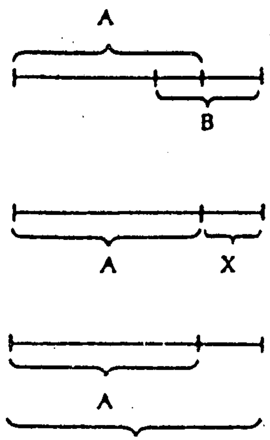


Figure 3.11.2
Here we have
A X = A⊕B, but X ≠ B

The pupils are allowed to think awhile and are then welcomed to give examples for or against. The results are unanimous: we cannot draw the conclusion that X = B, as Figure 3.11.2 gives witness to.

"Did Leibniz also think of that?" Yes, in a "scholium" (explanatory note) he asserts that "Proposition 10" cannot be reversed. He also gives other examples of non-reversible theorems in interval algebra.

Even if "all the truths of reason" do not quite fit into Leibniz' interval arithmetic, we must still honor him for having developed a new system of calculation, a system in which the symbols no longer represent numbers but rather intervals or sets of points.

3.11.2 *Some Applications*

"May we see an example of such interval arithmetic in practice?" Of course! Suppose that 100 people are asked about their knowledge of French and German: "How many consider that they have tolerable command of these languages, in the event of a trip to France or Germany or for reading a book?"

Suppose that 35 people considered that they knew French tolerably and 75 German. We now ask ourselves:

a) How many of the 100, as a minimum, are certain to command both French and German?

b) How many at most could command both French and German?

c) How many at most could speak neither language?

d) How many must, as a minimum, have a knowledge of at least one of the languages?

We let F be the set of people who know French, G the group of people who know German. The letter n will denote number, e.g. we will write

n (G) = 75, n (F) = 35

In order to answer question (a) we must separate G and F as much as possible; the overlapping part is to be minimized. We work forward to Figure 3.11.3 and find that

Min n (G and F) = 10

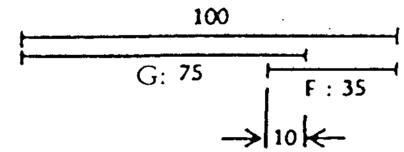


Figure 3.11.3
Min n(G and F) = 10

The number 10 can be obtained in different ways, e.g. as 35 - 25 or as 75 + 35 - 100 (the overlapping portion is left over in this subtraction).

In a similar manner we get the answers to the other questions:

b) Max n (G and F) = n (F) = 35, when F is included in G.
c) Max n (neither G nor F) = 25, when F is contained in G.
d) Min n (at least one lang.) = 75, when F is contained in G (Figure 3.11.4).

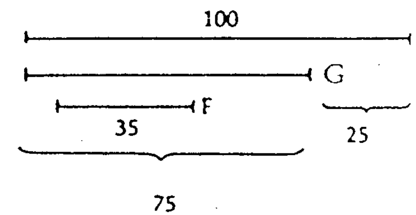


Figure 3.11.4
b) Max n(G and F) = 35
c) Max n (neither G nor F) = 25
d) Min n (at least 1 language) = 75

If we were to formulate similar problems concerning 3 or 4 languages, the difficulties would increase considerably. And yet abstract combinatorics came to be more and more concrete and manageable for 19th century mathematicians in England. We shall next turn to the most prominent of these mathematicians, George Boole.

3.11.3 *Boole (1815-1864)*

George Boole can be seen as the most prominent founder of abstract algebra. He was, like Leibniz, a genius at language (translating poems by Horatio at 12 years of age). Besides Latin and Greek, he interested himself in modern languages such as French, German, and Italian. Boole, who was more or less self-educated, was employed as a teacher at age 16 and started his own school in Lincoln, England, at age 20.

On the recommendation of a mathematician in London, A. de Morgan, Boole was appointed professor of mathematics at a newly started college in the town of Cork, Ireland, where he worked until his death in 1864. Boole achieved great success both as a lecturer and as a mathematician. His best works were an analysis of logic from a mathematical point of view (1847) and in particular "An Investigation of the Laws of Thought..." with applications in mathematical logic and probability theory (1854). The importance of his contributions was first recognized by Bertrand Russell (about 1910) who considered Boole's Investigation of the Laws of Thought to be the first truly pure mathematics.

When Boole was alive, Leibniz' essay on combinatorics had likely been completely forgotten. It seems improbable that Boole could have known of that work, partly because Leibniz' basic ideas were not even given attention in Germany, partly because Boole's concepts in many ways differed from Leibniz'. Boole, for example, used "addition" in another sense, letting a + a remain 2a and interested himself primarily in his new arithmetic's applications in logic. He investigated, for example, how such conjunctions as "both-and," "either-or," "if, then" and other expressions could be given arithmetic analogy. He formulated a number of axioms for his symbolic logic. We will here look at the set theory which Boole's followers, primarily John Venn (1834-1923), came to develop.

After serving as a clergyman and lecturer, Venn switched over to mathematics. From him we have the rather well-known Venn or set-diagrams which made set theory considerably more understandable and easily manageable than interval arithmetic. In 1864 W. S. Jevons (1835-1882) published the work "Pure Logic, or Qualitative Logic as Distinguished from Quantitative" and made a clear distinction between the concepts "a or b is true" (meaning a or b or both are true) and "either a or b is true" (one or the other is true). Venn accepted and used Jevons' system. We can put this in the terminology of set algebra in the following way: we let a Roman one, I, denote a set of something, for example 1000 people. We let A denote a subset of I (perhaps including the whole set), for example, those people who read American fiction, and let B be another subset, those who read British fiction.

We introduce further

A′ = the complementary set of A
= the set of people who do *not* read American literature = all those among the 1000 who are not members of A

and analogously B′,
as well as two operations ∪ and ∩

A ∪ B = "the union of A and B"
= the set of people who belong to A or B or both:
= the set of elements which belong to A *or* B

A ∩ B = "the intersection of A and B"
= the set of elements which belong to A *and* B
(i.e. to both A and B)

We can consider the forming of a complement as an operation which acts upon one set while union and intersection are operations which affect two sets. Figure 3.11.5 shows the graphical equivalents as Venn diagrams.

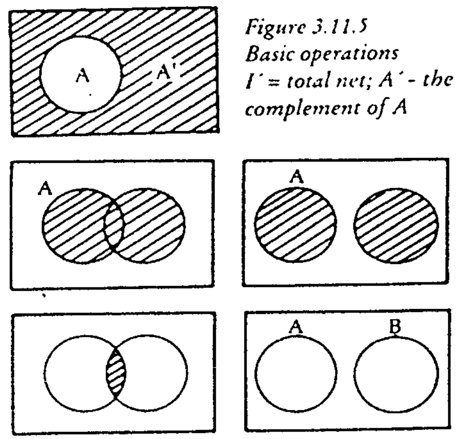


Figure 3.11.5
Basic operations
I′ = total net; A′ - the complement of A

A∪B

A∩B = Ø = empty set when A and I do not overlap

If A and B have no common elements, they are said to be separate or non-overlapping or disjoint. Their intersection then has no elements, which is usually expressed: the intersection = "the empty set" and one writes

$$A \cap B = \emptyset$$

Here it is interesting to compare this with a · b = 0 in arithmetic. Further examples of ground rules are:

$\emptyset \cup A = A$ $\quad \emptyset \cap A = \emptyset$ $\quad A \cup A' = I$

$I \cup A = I$ $\quad I \cap A = A$ $\quad A \cap A' = \emptyset$

and $(A')' = A$

Before going one with further development of the system of axioms we should, as in the classroom, first have a look at a concrete example of calculation with sets in a Venn diagram.

Suppose that the number of people in sets A and B above are 560 and 780 respectively and that 200 people belong to neither A nor B. Can we, from these figures, obtain other information? Can we, for example, determine that a certain number of people must have read both American and British literature, i.e. must belong to the intersection?

We draw a Venn diagram as in Figure 3.11.6. Obviously there are people in the intersection, in the lens-shaped overlapping area. But how many? There are different ways to proceed.

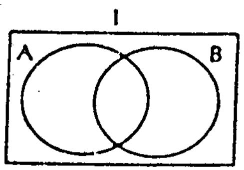


Figure 3.11.6
n(I) = 1000
n(A) = 560
n(B) = 780
n(neither A nor B) = 200

We can let x be the number of people in the intersection and then add up three separate sets which together make up the union:

$$(560 - x) + x + (780 - x) \text{ that is } 1340 - x.$$

This expression must be equal to $1000 - 200 = 800$, so we get the equation $1340 - x = 800$ and $x = 540$.

A direct method:

The number of people in the intersection is $(560 + 780) - 800 = 540$.

The line of thought is that adding 560 and 780 gives a double booking of the intersection amount. There are not so very many problem variations with 2 sets, but 3 sets give more. Of much greater interest, however, is set theory's structural foundation.

3.11.4 *Rules of Arithmetic*

With the aid of Venn diagrams, the class can investigate which rules should be included in the foundations of set theory. Which of the following ought to be considered "laws" in set theory?

$A \cup B = B \cup A$
$A \cap B = B \cap A$ commutative laws?

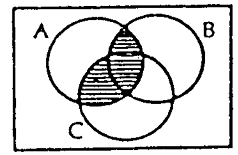
$(A \cup B) \cup C = A \cup (B \cup A)$
$(A \cap B) \cap C = A \cap (B \cap A)$ associative laws?

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ distributive law for intersection?

Corresponding laws apply in normal arithmetic if we let addition correspond to union and multiplication to intersection.

We find that all of these laws must be accepted. The last of them is illustrated in Figure 3.11.7a-b:

Figure 3.11.7a
$A \cap (B \cup C)$



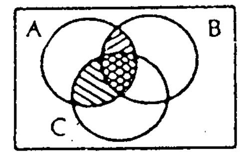


Figure 3.11.7b
//// $A \cap B$
\\\\\\ $A \cap C$

Might we expect that there is also a distributive law for union, i.e. that the following should apply?

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)?$$

We compare with normal algebra: is $5 + 2 \cdot 8 = (5 + 2) \cdot (5 + 8)$?

No!

We draw Venn diagrams for both the left and the right side of (*). See Figure 3.11.8.

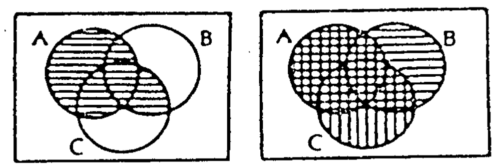


Figure 3.11.8
Three overlapping sets
$A \cup (B \cap C) = (A \cap B) \cup (A \cap C)$

What do we see in the figures? We see that if A, B and C overlap one another, then (*) is true!

It is not difficult to check that this result applies completely generally.

It turns out that the operations union and intersection ($\cup$ and $\cap$) participate entirely symmetrically in the system of axioms. They take on dual roles. For every general relation which is true there corresponds a dual relation which is obtained from the first by interchanging $\cup$ and $\cap$ everywhere, and interchanging I with $\emptyset$, while these are dual elements. For example, the relation

$$A \cap A' = \varnothing$$

transforms into the true relation

$$A \cup A' = I$$

More examples are given in the exercises. The basic dualities are, in summary, the following: For all A and B,

$$\begin{cases} A\cup B = B\cup A & (1a) \\ A\cap B = B\cap A & (1b) \end{cases}$$

$$\begin{cases} (A\cup B)\cup C = A\cup(B\cup C) & (2a) \\ (A\cap B)\cap C = A\cap(B\cap C) & (2b) \end{cases}$$

$$\begin{cases} A\cup(B\cap C) = (A\cup B)\cap(A\cup C) & (3a) \\ A\cap(B\cup C) = (A\cap B)\cup(A\cap C) & (3b) \end{cases}$$

$$\begin{cases} A\cup\varnothing = A & (4a) \\ A\cap I = A & (4b) \end{cases} \qquad \begin{cases} A\cup I = I & (5a) \\ A\cap\varnothing = \varnothing & (5b) \end{cases}$$

$$\begin{cases} A\cup A' = I & (6a) \\ A\cap A' = \varnothing & (6b) \end{cases}$$

We have thus come upon an example of duality even in an algebra, one of a more abstract nature than the duality we found in plane projective geometry (Section 3.10).

3.11.5 *Some Applications of Set Theory*

From set theory the step is not far to problems in probability and combinatorics.

For example, the basic set I could be the outcome of throwing 4 casts of a dice,

A might be the subset of outcomes containing at least one six
B might be the subset of outcomes containing at least one one, etc.
Then A′ would be the subset of outcomes with no sixes
B′ would be the subset of outcomes with no ones

and $A \cup B$ = the subset of outcomes which contains at least one six or at least one one

$A \cap B$ = the subset of outcomes which have both one or more sixes and one or more ones.

The number of possible outcomes, elements, in this basic set I is

$$6^4 = 1296.$$

We leave combinatorics and instead ask ourselves how Boole and other mathematicians went about formulating the outlines of a mathematical logic.

A, for example, could represent the statement "It is raining"
A′ would then mean the opposite of A (not-A): "It is not raining"
B could be another statement: "It is warm"
B′ would be its opposite: "It is not warm"
$A \cup B$ corresponds then to: "It is raining or warm"
$A \cap B$ corresponds to: "It is raining and warm"

From daily life we are familiar with logical statements of the type, "If the animal is a cod, then it is a fish." One can illustrate this statement using a Venn diagram (Figure 3.11.9): we let a C-circle represent the set of cods and an F-circle the set of all fish. The rectangle can represent the set of animals species. Since the C-circle lies within the F-circle, every point in C is also a point in F. That the C-set is contained in the F-set is usually written

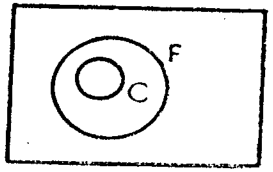


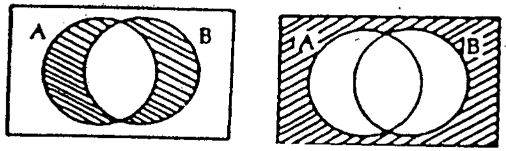
Figure 3.11.9
C ⊂ F: "If C, then F".

$$C \subset F.$$

Reversing "If C, then F" to "If F, then C" is, of course, wrong: if we have a fish, it need not be a cod. A point in the F-circle need not belong to the C-circle.

The illustration of the propositions of the type "If A, then B" with an A-set inside of a B-set is thus completely natural. Besides the expressions OR, AND and IF-THEN, the expressions EITHER-OR and NEITHER-NOR also have their graphical representations (Figure 10):

Either - or



Neither - nor

Figure 3.11.10 (overlapping sets)

3.11.6 *Logic and Truth*

It is interesting to follow the development of the "IF-THEN" relation in presentations of symbolic logic. At one point of time the implication, "If A, then B" or "A implies B" took on the following form: "A is false, or A is true, and in this case B is true." For example, we can reformulate the statement, "If the voltage in the line exceeds 100 volts, the transistor will be destroyed" to read: "The voltage in the wire is not more than 100 volts, or it is more than 100 volts and then the transistor will be destroyed." One can say more laconically: "The voltage in the line does not exceed 100 volts, or the transistor will be destroyed." The word "or" has here a broader meaning than the exclusive "either-or": nothing has actually been said about the transistor if the voltage is less than 100 volts, e.g. 99 volts. We have a tendency in everyday language to interpret 100 volts as the maximum safety limit for the transistor, but this is a generalization.

The laconic formulation of the implication can be represented by the expression

$$A' \vee B \quad \text{(not-A or B)}$$

where A stands for "the voltage exceeds 100 volts," and B stands for "the transistor is destroyed" (Figure 3.11.11).

The implication of "If A, then B" which in set theory's symbolic language is written $A \subset B$ (A is contained in B), has therefore been replaced by the expression $A' \vee B$.

In this version logicians went so far trying to simplify that they even allowed false if-statements to "imply" true then-statements. In *The World of Mathematics* (ed. James R. Newman), chapter XIII, section 4, Alfred Tarski gives the "implication"

"If 2 times 2 is 5, then New York is a large city"

as an example of a true statement.

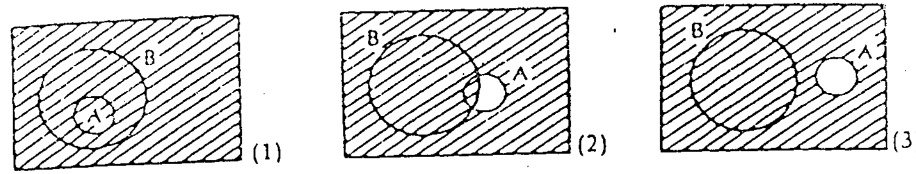


Figure 3.11.11

$A' \vee B$ in three cases as Venn diagram (shaded area)
A represents "The voltage exceeds 100 V"
B represents "The transistor is destroyed"
A' represents "The voltage does not exceed 100 V"
B' represents "The transistor is not destroyed"
Only the first case can illustrate the statement
"If the voltage exceeds 100 V then the transistor will be destroyed"

The motivation is, by analogy with the transistor example, that the statement reformulates to: "Not-P or Q is true," where P stands for "$2 \cdot 2 = 5$" and Q is "New York is a large city."

Even the following statements are considered to be meaningful:

(1) "If $2 \cdot 2 = 5$, then New York is a small city"
(2) "If $2 \cdot 2 = 4$, then New York is a large city"

Because one of the clauses in the reformulations

(1) "$2 \cdot 2$ is not 5 or New York is a small city"
(2) $2 \cdot 2$ is not 4 or New York is a large city"

is true (the first in (1), the second in (2)), the statements are accepted as meaningful. One need not be surprised to hear the following words by Alfred Tarski (from *The World of Mathematics*, Ch. XIII, 4, edited by James R. Newman): "The divergency in the usage of the phrase "if ..., then ..." in ordinary language and mathematical logic has been at the root of lengthy and even passionate discussions..."

> The two facts, that a false clause can imply any arbitrary clause and that a true clause is implied by any arbitrary clause, are sometimes referred to as the implication paradoxes. Yet they are not real "paradoxes." They express a discrepancy between the concept of material implication on the one hand and the concepts of conditional relation and consequence relation on the other. The stated discrepancy is remarkable. One can say that it has not been paid sufficient attention by the classics of modern logic.
>
> *(G. H. von Wright in "Logik, filosofi och Språk")*

In the simplified so-called material implication, exemplified above, the clauses can, as we have seen, completely lack any meaningful relation. But the material implication has been an effective tool when its purpose was to solve problems in logic with rational methods patterned along the lines of arithmetic. Here there are many opportunities in a higher class to take up discussion of logic, reality, and language as different modes of expression, and thereby bridge over into other school subjects.

A particularly meaningful theme is to train the pupils in understanding the difference between "If ..., then ..." and "If and only if ..., then ....." We ought to be aware of the meaning of necessary conditions, sufficient conditions, and conditions which are both necessary and sufficient. We will return to this subject in Section 6.2. See also Section 3.5.3.

The laws which were originally formulated for set theory can strangely enough be used in such divergent areas as logic, probability, and the theory of electric circuits. The algebra which presides over these areas was given the name Boolean Algebra.

3.11.7 Switching Networks

We shall conclude the chapter by looking at a few examples which give a hint as to how Boolean algebra can be applied in switch circuits.

In the theory of switch networks there exist only two values, 0 and 1. We define two operations, "addition" and "multiplication" with the tables

$$\text{(A)}\quad \begin{aligned} 0+0&=0\\ 0+1&=1\\ 1+0&=1\\ 1+1&=1 \end{aligned} \qquad\qquad \text{(M)}\quad \begin{aligned} 0\cdot 0&=0\\ 0\cdot 1&=0\\ 1\cdot 0&=0\\ 1\cdot 1&=1 \end{aligned}$$

The only unusual thing, as we see, is that $1 + 1 = 1$.

The value 0 is applied to an open switch or a non-glowing lamp. (Figure 3.11.12).

Summarizing, we can say that 0 corresponds to an open circuit and 1 corresponds to a closed circuit.

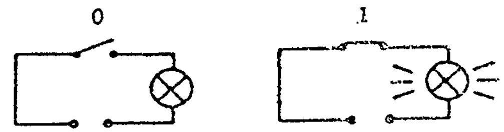


Figure 3.11.12

In the function table below we establish that (A) above corresponds to all possible situations with two switches in parallel, while (M) corresponds to two switches in series.

	Switch x	Switch y	Functional value f for the lamp
(A)	0	0	0
	0	1	1
	1	0	1
	1	1	1
(M)	0	0	0
	0	1	0
	1	0	0
	1	1	1

These tables are illustrated in Figure 3.11.13.

We can now step by step test the structure of switching network theory in relation to set algebra (we take only a few steps here):

For a switch value x, there corresponds the opposite switch value x′ such that

$x = 1$ is true whenever $x' = 0$ and

$x = 0$ is true whenever $x' = 1$

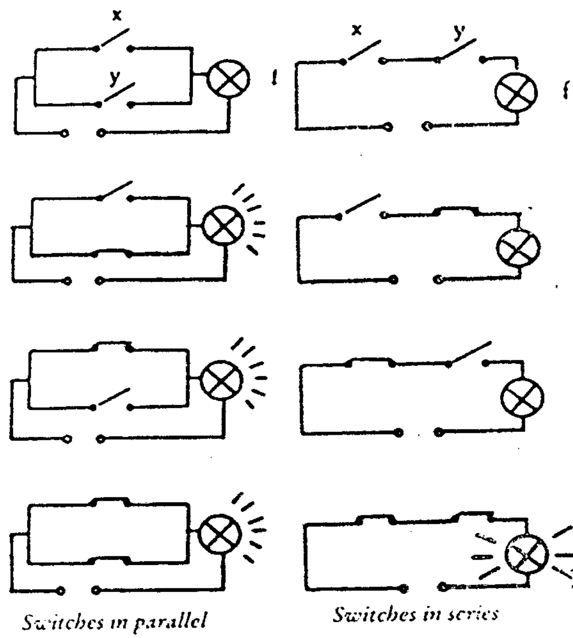


Figure 3.11.13

We note that

$$x + x' = 1 \quad \text{(always)}$$
$$x \cdot x' = 0 \quad \text{(always)}$$
$$(x')' = x \quad \text{(always)}$$

See Figure 3.11.14.

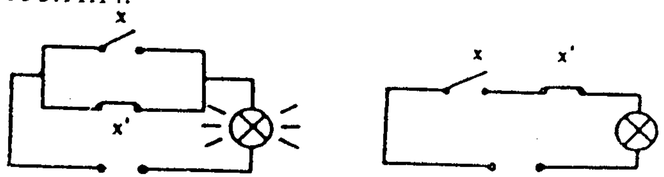


Figure 3.11.14
$x + x' = 1$ and $x \cdot x' = 0$ respectively. Illustrated here for the case $x = 0$.

Let us now ask: do the following laws hold:

(1) $x + y = y + x$?
$x \cdot y = y \cdot x$?

(2) $(x + y) + z = x + (y + z)$?
$(x \cdot y) \cdot z = x \cdot (y \cdot z)$?

(3) $x + (y \cdot z) = (x + y) \cdot (x + z)$?
$x \cdot (y + z) = (x \cdot y) + (x \cdot z)$?

It is easy to verify (1) and (2). That laws (3) also hold is taken up in one of the exercises below.

In this way one can discover that the axioms of Boolean algebra have their analogous counterparts in the theory of electrical circuits.

3.11.8 Exercises

1. In a class of 30 there were 25 who were positive toward Germany and 21 toward France as the country to visit on a school trip.
   a) Determine the maximum number of students who could be in favor of both countries.
   b) How many, at a maximum, could be against both countries?
   c) How many, at least, must be in favor of both countries?

2. In an opinion survey the people polled were asked among other things to answer yes or no to questions of the type: "Do you have confidence in politician A?" 80% showed confidence in A, 70% in B, and 60% in C, when the questions were asked separately one at a time. What per cent of the people asked, as a minimum, must then have confidence in all three of the politicians?

3. Draw two Venn diagrams with two overlapping subsets A and B. Show $A' \cup B'$ in the diagram and $(A \cap B)'$ in the other. Are these two sets identical?

   Investigate if this result holds even when A and B are non-overlapping sets and when A is wholly contained in B.

4. The same as Exercise 3 but with the sets $A' \cap B'$ and $(A \cup B)'$.

5. Exercise 3 and 4 illustrate de Morgan's laws:

$$(A \cup B)' = A' \cap B'$$
and
$$(A \cap B)' = A' \cup B'$$

   Show that these two identities are each other's duals (refer to Section 3.11.4).

6. Verify with the aid of an x - y - z table the identities

$$x + (y \cdot z) = (x + y) \cdot (x + z)$$
and
$$x \cdot (y + z) = (x \cdot y) + (x \cdot z)$$

(the distributive law for the operations + and · respectively, in switch circuits).

7. Show with the help of a function table that the switch circuit in figure (a) below is equivalent to the the simpler circuit in figure (b), that is, show that $f \equiv g$.

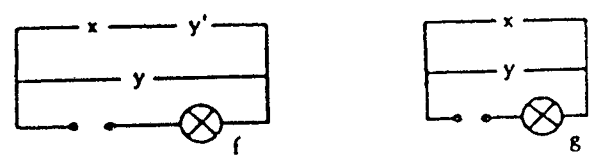


Figure a *Figure b*

3.12 Exercises in Concept Formation

3.12.1 From Galileo to Riemann

From Galileo's study of little balls rolling down the inclined plane, a red thread leads forward to our time and technology with moon landings, satellites and missiles. This thread goes via the method of calculation which Newton and Leibniz separately developed concerning derivatives.

Galileo wished to determine how far a ball rolls down a plane during a specified given time. In order to arrange the experiment methodically he began with an assumption about the ball's velocity in motion down the plane: that the velocity is proportional to the time. From this assumption Galileo succeeded in *calculating* the distance rolled. He could then arrange his experiments such that they gave him information concerning the correctness of his theory. He found complete confirmation that his intuition had been right.

Newton turned Galileo's approach around. At that time, thanks to Galileo, the rolling ball's distance was well known. Newton posed the reverse problem: how can we calculate the instantaneous velocity of the ball, if we know the rolling distance as a function of time?

Out of this question Newton formed the concept which nowadays is called the derivative and which is the mathematical tool for calculation of instantaneous velocity, among other things. This work was one of many famous achievements made by Newton at the age of 22-23 while Cambridge University was closed in 1665-66 due to the plague. It is an exceptionally important exercise in thinking to start from the well-known concept of average speed and think through the thoughts which lead up to the notion of instantaneous velocity, and more generally, to the derivative of a function, i.e. to the rate of change of a function.

From several simple examples (taken out of train timetables, diagrams, the t^2-expression for the ball's rolling distance, etc.) we soon arrive at the formula

$$\frac{f(b) - f(a)}{b - a}$$ for the average velocity of a function in an interval $a \leq x \leq b$

To get at the instantaneous velocity at $x = a$ we need to make b variable and let b both decrease toward a and increase toward a. The interesting thing is that we cannot simply put $b = a$ in the expression above. We would then get

$$\frac{f(a) - f(a)}{a - a} = \frac{0}{0}$$ which doesn't tell us anything.

We must carry out the passage to the limit $b \rightarrow a$ and investigate what happens to the quotient

$$\frac{f(b) - f(a)}{b - a}$$

It turns out that thanks to our knowledge of algebraic rules for factoring expressions, squaring and so on, we can derive a limiting value for the quotient above, as b goes toward a. (I must leave out this development and refer those readers who are not familiar with the concept of derivatives to a high school text.)

It is very interesting from a theory of knowledge point of view, that with the derived limiting value we are able to define what we mean by instantaneous velocity, a notion that we managed to grasp intuitively long before now.

It takes considerable time and requires much careful work before a heterogeneous class of 17-18 year old pupils can feel they have mastered the concept of derivative. It is interesting for them to experience how simple the actual arithmetic is in problem applications, compared with all the effort required to develop the concept.

3.12.2 Galileo

Let us return to Galileo and ask: how did Galileo calculate the rolling distances theoretically, before he went to experiment? He assumed — at least after a certain amount of "playing around" — that the velocity of a ball which starts from rest and begins rolling down a plane is proportional to the time, i.e. that

$v = k \cdot t$ where v = velocity
k = a constant
and t = time from start

How would he now calculate the accumulated distance rolled? Galileo began by assuming *uniform* motion, i.e. motion where the velocity is constant, say v_0. For this case the distance is of course $s = v_0 t$, that is, simply the product of velocity and time.

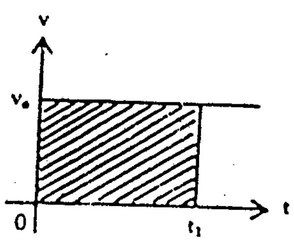


Figure 3.12.1

In a plot of velocity as a function of time (Figure 3.12.1), the identity $v = v_0$ is a straight line parallel with the t-axis. The value of the distance up to a time $t = t_1$ is $v_0 t_1$. In the plot this value corresponds to the area of the shaded rectangle. For uniform motion then, the accumulated distance is proportional to the area of the corresponding rectangle in the v-t-plot.

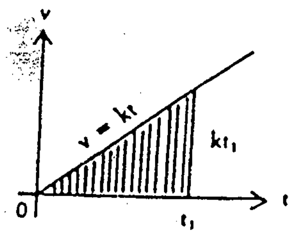


Figure 3.12.2

Galileo assumed now that an analogous relation holds even for the case where $v = kt$, i.e. he assumed that the distance is proportional to a triangular area (Figure 3.12.2) enclosed by the lines $v = kt$, $t = t_1$ and the t-axis.

This area is easily calculated to be:

$$A = \frac{kt_1 \cdot t_1}{2}$$

or

$$A = \frac{kt_1^2}{2}$$

The distance rolled should thus be proportional to the square of the time, which Galileo, as we know, confirmed through his experiments.

Was it audacious of Galileo to assume that the area calculation would give correct results even for accelerated motion? Hardly. We can imagine the linearly increasing speed approximated by a staircase function, where velocity increases in small jumps and is constant during the sub-intervals of time. With such a velocity the distance would be the sum of a number of products (Figure 3.12.3), and it is not hard to imagine passing toward the limiting velocity function

$$v = kt$$

in such a way that one lets the length of the sub-intervals of time go toward zero. The staircase function then approaches the straight line $v = kt$ everywhere, and the sum of the sub-areas nears, as closely as we like, the triangle's area

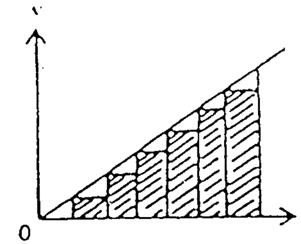


Figure 3.12.3

3.12.3 The Riemann Integral

Ought not this method of approximation with a staircase function work with other velocity functions than the linear? If the answer is yes, we would command a method of calculating distances for any given velocity function.

What must we require of the velocity function, when it is more "difficult" than a linear function? We note the requirement "positive continuous function," which primarily means that the velocity curve must be in one unbroken piece.

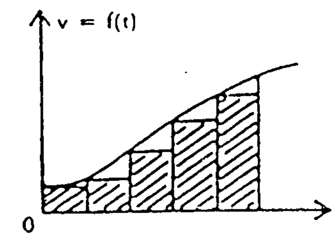


Figure 3.12.4

The graphical picture might then look as shown in Figure 3.12.4, where an approximating staircase function is drawn in under the curve. In order to do the approximation more accurately, we also draw in a staircase function with oversized values (Figure 3.12.5).

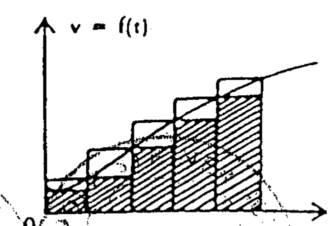


Figure 3.12.5

We can now begin to develop the integral which Bernhard Riemann (1826-66) gave form to in a dissertation the year 1850 (at the age of 24!).

To begin with we subdivide the interval $a \leq t \leq b$, over which the function $f(t)$ is continuous, into n sub-intervals

$\Delta t_1, \Delta t_2, \Delta t_3, \ldots, \Delta t_n,$ where

$\Delta t_1 = t_1 - a, \Delta t_2 = t_2 - t_1, \ldots, \Delta t_n = b - t_{n-1}.$

We denote this subdivision as A.

According to a well-known theorem of analysis, and intuitively understandable to the pupils, a continuous function always has a largest value and a smallest value on a closed interval. We can therefore introduce the nomenclature

M_k = max f(t) on the sub-interval Δt_k
m_k = min f(t) on the sub-interval Δt_k

We imagine now that f (t) represents a velocity function. The distance during the time interval t_k can then be approximated by

the product $M_k \Delta t_k$ (too large a value)
and by the product $m_k \Delta t_k$ (too small a value).

The accumulated distance for the whole time interval $a \le t \le b$ can then be approximated by the slightly high *upper sum*

$S_A = M_1\Delta t_1 + M_2\Delta t_2 + \ldots + M_n\Delta t_n$ and by the slightly

lower sum $s_A = m_1\Delta t_1 + m_2\Delta t_2 + \ldots + m_n\Delta t_n$.

The question is now how dependent these values are on the particular subdivision of the interval, and what happens to them if we let the length of the sub-intervals approach zero. These two questions are interwoven with each other, as we shall see. We shall investigate them step by step and come to the concept of the Riemann integral.

1. What happens to the upper and lower sums if one adds subdividing points and thus gets more but smaller intervals?
Answer: The upper sum cannot increase, but may decrease;
the lower sum cannot decrease, but may increase.

2. Can any lower sum be larger than any upper sum?
Intuitively the answer is easy: no.
There is an elegant proof of this (given by the Frenchman G. Darboux in the 1880's):

We are to prove that if A and B are two arbitrarily chosen subdivisions of the interval, then

$$s_A \le S_B \qquad (1)$$

In order to prove this we let C be the subdivision of the interval which the A-points and the B-points together create (Figure 3.12.6). Then, by the conclusion of step 1. above,

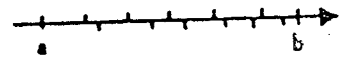


Figure 3.12.6
Marks pointing upward: A
Marks pointing downward: B
All marks: C

$$s_A \le s_C \le S_C \le S_B$$

from which $s_A \le S_B$.

(The inequality $s_C \le S_C$ we have understood earlier.)

3. From (1) it follows that the set of all values given by lower sums lies to the left, on the real number axis, of the set of all values given by upper sums. The question now is: have these two sets of numbers a common limit (upper limit for the lower sums, lower limit for the upper sums) or are they separated by an interval? (Figure 3.12.7).

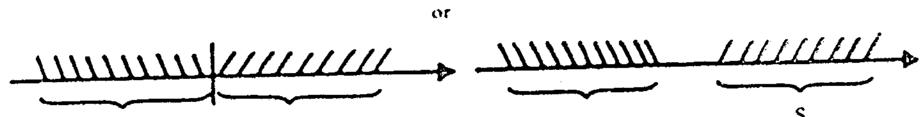


Figure 3.12.7

4. Darboux' version of the Riemann integral now is: If both value sets have a common limit, we define that number as the integral of f (t) on the interval $a \le t \le b$ and denote it by

$$\int_a^b f(t)\, dt$$

If the sets are separated by an interval, the function f (t) is said to be non-integrable on the interval $a \le t \le b$.

5. We shall now check whether the continuity of f (t) insures the existence of the integral. Do the lower sums and upper sums have a common limit? It is sufficient to show that the difference S-s can be made arbitrarily small, if the interval subdivision is made finer and finer, i.e., as the sub-interval length goes toward 0.

We get

$$S - s = (M_1 - m_1) \cdot \Delta t_1 + (M_2 - m_2) \cdot \Delta t_2 + \ldots + (M_m - m_n) \cdot \Delta t_n \quad (2)$$

We must here recall a theorem (on so-called uniform continuity) that *all* the differences $M_k - m_k$ can be made smaller than an arbitrarily small number Σ, by insuring that all Δt_k are shorter in length than a sufficiently small number δ.

Then this holds true:

$$S - s \leq \Sigma (\Delta t_1 + \Delta t_2 + \ldots + \Delta t_n) = \Sigma (b - a).$$

Since Σ can be chosen arbitrarily small, S-s can be made arbitrarily small, and we have insured the integral's existence.

6. One of the results of step 5 is the extremely important conclusion that the reasoning there holds independent of how the interval subdivision is done; the only thing that matters is that the length of the sub-intervals approaches zero.

The value of the integral, as a common limit for the lower and upper sums as max Δt_k approaches zero, is thus independent of how the sequence of finer and finer subdivisions is made.

7. If f(x) changes sign, one analyzes separately these sub-intervals determined by the zero points. The construction may then be carried through as before.

We now know that the integral of a continuous function *exists,* but how can it be calculated? We have little use for knowing of its existence if we cannot calculate its value!

This situation is rather typical in the area of mathematics dealing with limits: first comes proof of existence ("the limit exists") and after that calculating its value. (In the derivation of the derivative these occurred simultaneously: the existence of the limit value was shown along with its calculation.)

3.12.4 How the Riemann Integral is Calculated

We study the function

$$I(x) = \int_a^x f(t)dt \qquad \text{for } a \leq x \leq b$$

and ask ourselves: does the derivative of this function exist? To investigate this we form the difference quotient (from our knowledge of the definition of the derivative):

$$\frac{I(x+h) - I(x)}{h} = \frac{1}{h}\left(\int_a^{x+h} f(t)dt - \int_a^x f(t)dt\right).$$

From the construction of the integral, it follows directly that the difference on the right hand side is equal to the integral over the interval $x \leq t \leq x + h$, and we get

$$\frac{\Delta I}{h} = \frac{I(x+h) - I(x)}{h} = \frac{1}{h}\left(\int_x^{x+h} f(t)dt\right).$$

Let m_h and M_h denote the minimum f (t) and maximum f (t) over the interval $x \leq t \leq x + h$. Then the following inequalities hold:

$$\frac{1}{h} \cdot m_h \cdot h \leq \frac{\Delta I}{h} \leq \frac{1}{h} \cdot M_h \cdot h$$

that is

$$m_h \leq \frac{\Delta I}{h} \leq M_h$$

Now when $h \to 0$, both m_h and M_h approach the value f (x), since f (x) is assumed to be continuous.

We therefore let h go toward zero and find that the limit of $\frac{\Delta I}{h}$ exists and is equal to f (x).

In other words: the function I (x) has a derivative (and thereby is also continuous):

$$I'(x) = f(x) \qquad \text{or} \qquad \frac{d}{dx}\left\{\int_a^x f(t)\,dt\right\} = f(x)\,^{1} \quad (3)$$

---

[1] Right-derivative or left-derivative at the endpoints x = a and s = b respectively.

A beautiful relation between integral and derivative!

But still we do not know how to *calculate* an integral!

Assume that we have found a function F (x) such that its derivative is f (x):

$$F'(x) = f(x)$$

Example: If $f(x) = x^2$ we can note that $F(x) = \frac{x^3}{3}$.

We call F (x) a *primitive* function of f (x) (a source function for f (x) in the sense that f (x) is the derivative of F (x)).

We then are in the situation that our function F (x) and the integral I (x) both are primitive functions of f (x); both have the same derivative, f (x). Here we refer to "The Fundamental Theorem of Integrals," which says that

if $I'(x) = F'(x)$ then $I(x) = F(x) + \text{a constant}$.

Popularly interpreted: if two trains, the one following the other, always keep the same speed, then the distance between them remains constant.

According to this theorem we have

$$I(x) = F(x) + C \quad (C = \text{constant})$$

$x = b$ gives us $I(b) = \int_a^b f(t)\,dt = F(b) + C$ (4)

$x = a$ gives us $\int_a^a f(t)\,dt = 0 = F(a) + C$ (5)

Subtracting (4) minus (5) now gives us

$$\int_a^b f(f)\,dt = F(b) - F(a)$$

and we have found the formula for calculating an integral.

If we continue with our earlier example, $f(x) = x^2$, and let a and b be 1 and 4 respectively, we get

$$\int_1^4 x^2 dx = F(4) - F(1) \quad \text{where} \quad F(x) = \frac{x^3}{3}.$$

The value of the integral becomes $\frac{4^3}{3} - \frac{1^3}{3} = \frac{64 - 1}{3} = 21$.

3.12.5 *So Much "Theory"?*

Are not sections 3.12.3 and 3.12.4 an overambitious load on the students? If one knows the class and makes this judgment, then one might let construction of the Riemann integral be a voluntary chosen extra work project. But isn't it so "beautiful" that it is worth the time required? And doesn't it give valuable *exercise,* both in abstraction and ability to hold several trains of thought together?

The calculation F (b) - F (a) does not require any particular thought when simpler functions are concerned. The calculation becomes a routine procedure. Here, as in the section on derivatives, the formula is very simple to use compared with the thought behind it.

3.12.6 *A Word about Karl Weierstrass*

We have seen that *all* functions which are continuous on a closed interval are integrable.

What about derivatives: have all continuous functions also a derivative? Are there functions which are continuous on an interval $a \le x \le b$ but which somewhere lack a derivative?

We only have to look at Figure 3.12.8 to see the plot of a function which is continuous but lacks a derivative at four places, at the points x_1, x_2, x_3 and x_4.

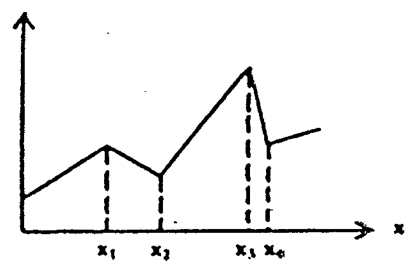


Figure 3.12.8

Can a continuous function lack a derivative everywhere? It seems to be completely impossible. It therefore aroused enormous surprise when Karl Weierstrass (1815-1897) presented a continuous function which does not have a derivative for any value of the variable.

The curve of such a function lacks a tangent everywhere, i.e. at every point it lacks a direction in spite of being continuous (unbroken).

One example of a function with this property is the sum of the infinite series

$$f(x) = \frac{\sin 3x}{2} + \frac{\sin 9x}{4} + \frac{\sin 27x}{8} + \ldots + \frac{\sin 3^n x}{2^n}$$

After Weierstrass other mathematicians have contributed new examples of continuous functions which everywhere lack derivative.

Weierstrass' results are a good example of how mathematics sometimes shows that "the impossible might be possible."

As another example of this, one can study the research results of Georg Cantor (1845-1918), who showed that there exist many different degrees of "infinitely many."

3.12.7 A Little on Georg Cantor's Research

Georg Cantor was a pioneer who developed set theory in the true sense of the term. What we today even in grade school call "set theory" (as part of the 'new math') is by and large completely apart from the area in which the real problems in set theory have their beginnings. In the school's "set theory" one speaks namely about problems involving *finite* sets. Cantor took on the study of *infinite* sets.

One of his very first questions was: are there greater, "bigger" infinities than the infinity we associate with the natural numbers 1, 2, 3 ...?

Let us call this set N and the set of all integers (N plus zero plus the negative integers) H. Is H a larger set than N? Cantor introduced a notion of size which meant that two sets A and B (containing elements a and b respectively) are equally large (equivalent in size) if the a- and b-elements can be put together in pairs, either finitely or infinitely in number. For example, the set of 6 oranges is equally large as the set of 6 apples. (It is such examples, among other things, which some school children are put to work on at a much too early age.)

And now returning to our question: are N and H equally large? It is quite clear that the set N is contained in H. We might perhaps say that it makes up about "half the amount." But maybe the N-elements can be put into pairs with the numbers in H. For example, the set of all numbers counting by "tens", 10, 20, 30, ... is equivalent in size to N, in spite of the fact that the terms make up only a fraction of N: we can namely form the following series of pairs, which runs through both sets equally quickly:

1 - 10
2 - 20
3 - 30
etc.

Two infinite sets can thus be equally large or equivalent even though the one is a part of the other. Here it is possible that the parts are as large (in the Cantor sense) as the whole.

Now, can the integers of H be put into pairs with the natural numbers? Yes, we can construct the following table:

H	N
0	1
1	2
-1	3
2	4
-2	5
3	6
3	7

etc.

H is therefore not greater in size than the set N; H and N are equivalent. We say that H is *countable* — forming pairs with the numbers of N is actually counting.

Is the set of all corner points (lattice points) in an xy-coordinate system countable? (Figure 3.12.9) Here we have an infinite number of horizontal lines and each contains infinitely many points. Is this set countable?

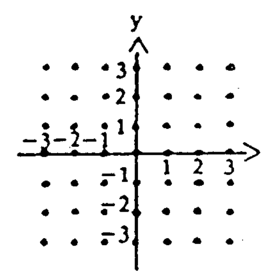


Figure 3.12.9

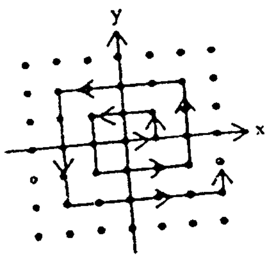


Figure 3.12.10

The answer is easily found: we begin at the origin (point 1) and number the points in an endless spiral outward (Figure 3.12.10). Each lattice point will get counted exactly once. Insight tells us now that the lattice points can be numbered so their set is equivalent to N.

Cantor went on to ask: can all the numbers between 0 and 1 be numbered, i.e. are they countable? They can be illustrated by the interval I of points between 0 and 1 on the x-axis. By way of introduction we can establish that any sub-interval of I, *however small*, must have the same size as I! Take for example, a fifth of the interval and place this one fifth-interval I´ perspectively opposite I, as shown in Figure 3.12.11. Each point x on I has a perspective point x´ on I´. We see that the points on I´ can in this manner be linked in pairs with the points on I. Their equivalence in size is then a consequence of the perspectiveness between I and I´ relative to the center A (Figure 3.12.12).

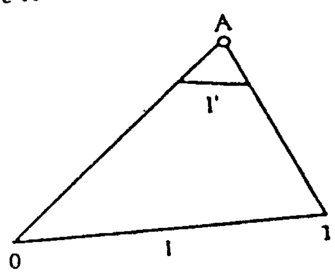


Figure 3.12.11

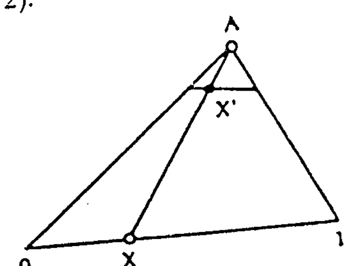


Figure 3.12.12

But now to I: are its points countable?

Let us with Cantor assume, *on a trial basis*, that I's numbers are countable. This means that the numbers could be written down on an infinitely long list, which would contain *all* the numbers between 0 and 1. Then, during the list's "printout," every number between 0 and 1 must sooner or later come out on the list. The numbers between 0 and 1 can be expressed as decimal fractions of the type:

0.abcd...,

where the letters stand for digits. In order to avoid double accounting of fractional numbers we agree that decimal fractions such as

0.34699999... (where all decimals after a certain place are nines)

will be written as in principle completed fractions with infinitely many zeros, in our example

0.34700000...

This is entirely correct, since 0.9999... must be assigned the value 1.

Let us now assume, for example, that the numbers in the beginning of the list look as follows:

number no. 1	0. 3 5 6 4 9 2 2 5 ...
number no. 2	0. 2 0 0 1 7 1 8 0 ...
number no. 3	0. 0 1 5 8 7 6 2 1 ...
number no. 4	0. 7 7 3 8 9 2 8 5 ...
etc.	

Cantor now shows that we can form a new number x (or even several), which cannot possibly be included in the endless list which has been given us and which claims to contain all the numbers between 0 and 1. He forms, for example,

x = 0. 4 1 3 7 ...

according to the following principle:

the 1st decimal in x	shall be different from the 1st decimal in the 1st number of the list (here 3)
the 2nd decimal in x	shall be different from the 2nd decimal in the 2nd number in the list (0 above)
the 3rd decimal in x	must differ from the 3rd decimal in the 3rd number in the list (5 in our list)

and so on.

We can choose the first four decimals to be, for example, 4, 1, 3, and 7. In continuing th nth decimal of x is to be different from the nth decimal in the nth number in the list

Number 1	0.	(3)	5	6	4	9	2	2	5...
Number 2	0.	2	(0)	0	1	7	1	8	9...
Number 3	0.	0	1	(5)	8	7	6	2	1...
Number 4	0.	7	7	3	(8)	9	2	8	5...

....

Through this Cantor insures that x cannot possibly be the same as any number in the list. To be on the safe side we must avoid giving x an infinite row of nines (if this were the case an x would appear to be different from the number 0.124000000... in the list but actually have the same value as that number, and thus x would be included in the list). Avoiding an infinite row of nines is no problem since for each decimal position we have 8 other digits than 9 from which to choose.

The assumption that all the numbers between 0 and 1 could be listed in an infinitely long list leads therefore to the contradiction that the list would not include all the numbers. This contradiction shows that the set of numbers from 0 to 1 *is not countable*. This set is "larger" than the set N. Cantor called this greater size the continuum.

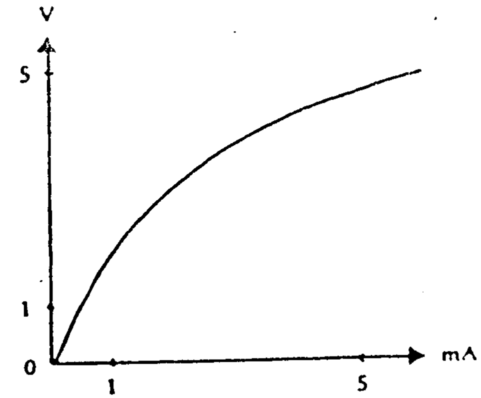
This result concerning the real numbers was one of Cantor's first discoveries. In an impressively energetic and persevering manner Cantor continued his journey of discovery through the realm of infinity. The work required much mental energy and alertness and took its toll on Cantor's health. On top of this his methods received strong criticism from some mathematicians. The strenuous intellectual work as well as opposition from authorities in the field eventually broke Cantor, and he was obliged to seek help at a mental hospital on several occasions.

When it was later shown, not least through Bertrand Russell's work, that set theory, as it had developed, led to paradoxes of a difficult nature, efforts were aimed at introducing greater stringency into the concept of set. Cantor's contributions maintained their great importance, and it is no exaggeration to see him as one of the boldest and most pioneering of all mathematicians.

3.12.8 Exercises

1. What average velocity does the Intercity train "Tiziano" have between Hamburg (departure 7.45) and Hannover (arrival 9.08)? The distance is 178 km.

2. With what average rate, expressed in volts/milliampère (V/mA), does the voltage in the diagram below increase, as the current increases from 1.1 mA to 5.5 mA?



Exercise 2

3. A ball starts from rest and rolls down a plane. The slope of the plane is such that the function for the distance rolled is

$$s(t) = 1.6t^2$$

where t is measured in seconds and s in meters. Calculate the ball's average speed during the time from t = 2 to t = 4 · 5.

4. Toward what limit does the expression

$$\frac{(t + h)^2 - t^2}{h}$$

go, as h goes toward zero?

First rewrite the numerator, and then show that the h in the denominator can be factored out and the fraction reduced.

5. The result in Exercise 4 is the limit of the average velocity during the time interval from t to t + h, for motion whose distance as a function of t is t^2.

More generally: the result of Exercise 4 is called the derivative of the function t^2. Try to derive in an analogous manner the derivative of the functions t^3 and t^4 respectively, and generally of the function t^n, where n is a natural number.

6. Show that if f (x) is continuous on an interval $a \leq x \leq b$ and

$$\int_a^b f(x)^2 dx = 0,$$

then it must be true that $f(x) \equiv 0$, i.e., $f(x) = 0$ for all values of x.

7. Calculate

a) $\int_0^1 (x - x^2)\, dx$

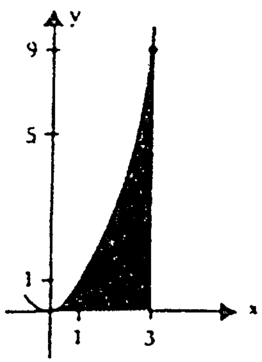


Figure 3.12.13

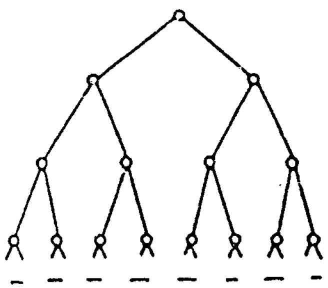


Figure 3.12.14

b) the area in the xy-coordinate system (with 1 cm scale units) bounded by the x-axis, the parabola $y = x^2$ and the line $x = 3$ (Figure 3.12.13)

8. Show that the set of square numbers,

$$1, 4, 9, 16, \ldots$$

is equivalent in size with the set of natural numbers.

9. Show that the set of rational numbers a / b between 0 and 1 is countable, i.e. has the same size as the set of natural numbers.

10. Figure 3.12.14 shows the start of an endlessly dividing "street network." Each road divides into two roads. Show that the infinite set of separate roads contained in this network is not countable. (None of the roads cross one another.)

4

MATHEMATICS AS A FIELD OF PRACTICE FOR THINKING

4.1 To Achieve Sureness in Thinking

Participants in mathematics courses for adults have sometimes said such things as "Mathematics can actually be interesting," "I was afraid I wouldn't be able to keep up in mathematics," "I can't remember that we worked with this kind of arithmetic in school," or "Mathematics was boring in school."

Perhaps teaching is considerably better nowadays than when we sat behind school desks. But the pupils' relation to mathematics is certain to be just as strongly individual now as then. There are students who like the subject, and there are others who find it painful. Quite apart from their attitude toward mathematics many students make an interesting comment when they have just conquered a difficulty in the subject, small or large: "It was really difficult until I understood it, but then it was nothing!"

Later we will go into the psychological aspects of the fact that problems are difficult until one understands them, but for the moment we direct our attention to another phenomenon which teachers may find in virtually all of their pupils: their joy when they have understood something, or even more so, when they have discovered something and best of all, when they have discovered something all on their own.

It is important to give the pupils free space in which they can think on their own. Some pupils gain a very strong motivation to work, when they just once solve a problem without help. "Don't tell us," "wait a minute," "no, I don't want any help" and other such expressions bear witness to the fact that many students wish to test their own abilities in mathematics (just as in other subjects)

This is not a question of achievement-oriented pupils, super-ambitious types, but rather of completely ordinary children and youth who want to keep watch over their own inner workings.

An experience which does not so easily come to expression in the classroom but which I would place highest of all, is the feeling of sureness, when one has attentively, consciously solved a problem. Sureness comes the moment we see that the method we have chosen will lead to the goal. It is not dependent on those calculations or manipulations we might need to do in order to get the answer. It is based on the experience of having found a way to reach the goal and of knowing that one can follow a thought process just as surely as one can walk down a road.

There are some interesting observations on how the experience of sureness can come to be present just as suddenly as when lightning strikes. In *The World of Mathematics* (edited by James R. Newman, New York 1956) we may read personal descriptions of this experience by the French mathematician Henri Poincaré (1854-1912). Poincaré had wrestled with the problems in the theory of functions for a time while in Caen, but left town and mathematical work to participate in a geological expedition:

> The changes of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions, were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty.

On a later occasion, after unsuccessful efforts with a problem in algebra:

> One morning, walking on the bluff, the idea came to me, with just the same characteristics of brevity, suddenness and immediate certainty, that...

And still another time, when the solution popped up while Poincaré strolled along a street:

> I did not try to go deep into it immediately, and only after my service did I again take up the question. I had all the elements and had only to arrange them and put them together.
>
> (*The World of Mathematics*, Chapter XVIII, section 2.)

I have quoted Poincaré in considerable detail not because we can expect pupils to have such marked experiences in school. But the same quality of knowing, even in a considerably humbler form, *can* be experienced in elementary mathematics. That the solution often comes suddenly, unexpectedly, is an experience shared by many inventors. But even in those cases where the sureness comes gradually, it is of the very greatest value as an inner experience. It appears most valuable when we succeed in finding the solution to a problem all by ourselves. But even when the sureness slowly grows while we follow another person's reasoning, for example in a proof of the Pythagorean theorem, it is an experience of inner clarity and control.

A person who actively experiences the soundness of the proof of the Pythagorean theorem *knows* that the theorem is true. This inner conquest cannot be taken away from him. It is not the case that we believe the Pythagorean theorem because it has been presented to us by mathematicians or because a teacher has explained the proof to us: we know it is true the moment we can grasp it with our own thought.

Mathematics is therefore a practical field which, despite the dependence of pupils upon their teacher, serves to free them from bonds to authority.

The teacher can of course encourage or restrain such a process of independence. The more the student is allowed to orient himself within a subject, the better. For this reason the students ought, at least in the beginning of each new topic, to be given opportunity to gain first hand experience with the arithmetical or geometrical subject materials which belong to the particular section of study.

The opportunity to try new materials is, as we know, exciting for most people in subjects such as chemistry and physics. What is more fascinating than trying out a new apparatus, simple though it may be? To test and to seek are ways of getting answers to questions which one poses. It is part of our instinct for knowledge to behave in this manner, to go on scouting expeditions hoping if possible to make discoveries. Big or little — it does not matter.

In arithmetic I have given examples of how students can go "scouting" and testing with numbers which they themselves have chosen and arranged (for example, Section 3.5.2). In geometry everyone can draw figures and look for what possible relationships can be uncovered in the figures, both in their own and in their classmates'. This element of looking, seeking, appeals to the individual and awakens his or her interest to see what others have come up with and to let others share in one's own results. In short experimenting and searching are a phase in which individualized activity goes hand in hand with social activity.

A truly stimulating tension occurs when one group of pupils have found one result, and another group reports the opposite results. No one need feel beaten when it is seen later which group has made the right judgment, because all the experiences were used in the search for knowledge.

It also gets exciting when we have the choice of proving a supposed lawful relation or of looking for counter-examples. Which horse do we bet on? A proof is often demanding, while a single counter-example is sufficient to throw out a theory we have. The art sometimes lies in being able to find a counter-example. If the new examples instead confirm the theory, shall we then change horses and try to prove the theory?

Once at the end of a lesson where we had treated the theorem that two triangles are similar when their corresponding angles are equal, I gave the following homework assignment:

Investigate whether or not the theorem applies to other figures than triangles, i.e. to polygons with four or more sides. You might draw four- or five-sided polygons and feel your way forward. The angles A, B, C, etc., in the one polygon must be the same size as the angles A´, B´, C´, etc. in the other polygon. The question is: are the polygons then similar? I assured myself that the students had understood the task and was then curious to see what they would come up with the following morning.

A large group in the class had found that the theorem for triangles was true even for polygons. One boy who usually was not particularly outspoken in class said that the theorem did not apply to other polygons than triangles. A smaller group of pupils had not come to any conclusion. The pupils in the large group felt very sure of themselves. Underneath that obviously lay a sense of belonging to the majority. One could note their attitude of superiority toward the boy who claimed the opposite. It was an experience to see their reactions, those in the majority, when the boy went up to the blackboard and drew differently shaped

polygons approximately as in Figure 4.1.1 below. (The simpler counter-example of the square and rectangle even this lad had missed.)

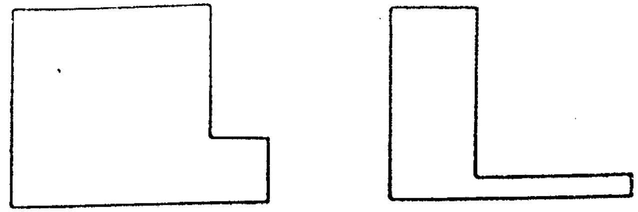


Figure 4.1.1

4.2 The Formal

The more time we use for heuristic, seeking out kind of work, the better. But to training in mathematics also belongs the learning of standard methods and becoming capable of using them and knowing when to use them. It would be an unhappy situation if students should begin searching for a new solution to a problem which they ought to recognize from previous examples. The inexperienced person often takes a long path to get to a solution where the person with experience only needs to make use of a simple idea to re-make the problem in terms of known technique. The risk here is that the mathematics disappears, as routine makes its entry, especially if strong emphasis is placed on memorized formulas or formula reference tables. At higher levels in school many students become conscious of the fact that calculation and mathematics are not identical occupations. They know that calculation can be done by machines. Even further from mathematics lies formalism, the stenographic dress of mathematics. How this and that are written is in reality a matter of convention. That the set of real numbers between 0 and 1 can be written $0 < x < 1$ or E $(x: 0 < x < 1)$ or (o;1) corresponds approximately to the fact that the symbol for four can be written IV or 4. It was a bad pedagogical mistake of "the new mathematics" to put so much emphasis on the rote learning of such things of convention. Through this, teaching took on the character of authority; the pupils became dependent on recipes just as the inexperienced person in cooking does.

Particularly fateful does this kind of teaching become if a number of students want quick results and get support for a question such as: "Tell us how we're supposed to do it" or, as sometimes occurs in physics experiments, "Why do all this experimenting? Tell us now how it's supposed to be." These students have momentarily only results before their eyes and do not want to take the trouble of seeking it out. That comfort can tempt pupils into making the book or their teacher into an authority was once demonstrated by the argument of a high school class to their teacher: "Why do you go through all these proofs? We believe you anyway."

And yet there may lie something admirable in such a student expression, a sound protest against a monotonous going through of proof after proof, often with the good intention of achieving stringency. Euclid's long list of definitions, theorems, and proofs (a pattern which repeats in many mathematical presentations) is not any example of *teaching* excellence. In school the theorems can easily become ends in themselves, just as much as formality. The fight against formalism must naturally not be allowed to lead to a fight against the requirement for logic in reasoning, points out the Russian mathematician A.J. Chintschin in an essay on formalism in teaching mathematics in school. In the same spirit are the words of the American Morris Kline:

> To teach thinking we must let the students think, let the students build up the results and proofs even if incorrect. Let them learn also to judge correctness for themselves. Let's not push facts down students' throats. We are not packing articles in a trunk. This type of teaching dulls minds rather than sharpens them.
>
> (From "Why Johnny Can't Add: The Failure of the New Math," 1973.)

Many who have studied the literature in mathematics or in other subjects, or who have listened to lectures, have noticed that passive, receptive study is tiring. Of course there are captivating books and lecturers, but sooner or later one longs to grapple with a problem on his own. It feels just as necessary as breathing out after a long breath in.

We are not especially creative when we with our thoughts follow after a train of thought laid down by another. School is unfortunately built entirely too much upon such "after thought." Mathematics in particular, where the study of proofs of propositions and theorems places upon us a thinking-after which can negatively effect our fantasy and ability to come up with ideas.

It is important for the mathematics teacher to have the personal experience that the deductive side of mathematics belongs to the synthesizing, concluding phase of problem-solving. It is preceded by an analytical, inductive phase where we seek after ideas, simplifications, associations to previous experience, etc., in order to get at the problem. Yet very first of all we must acquaint ourselves with the problem, listen to its different sides, try to put it into some context or familiar perspective. These phases of thinking we might call preparatory thinking, a "prethought" as opposed to "after-thought." We maintain an attitude in this stage much like an architect trying out various sketches of a building, which must meet certain specifications. It is interesting to experience again and again that such "pre-thinking" does not seem too tiring. Often it feels stimulating, it pushes away any possible tiredness, naturally enough because we are creative while we are doing it. Early on in the lives of children, we can develop a sense of this kind of thinking without forcing upon them any precocious intellectuality. Here, too, we include the art of making up and guessing riddles, of finding the right word, and other forms of play which emphasize thought. The students' appetite for "cracking" problems, such a valuable resource in teaching mathematics, must be awakened (but carefully!) while they still have the natural desire to seek out answers themselves.

4.3 Gaining Confidence in Thinking

Even those who do not solve many problems on their own in mathematics will come, sooner or later, to have *one* important experience: learning to be careful with what we happen to have in the way of subjective ideas, our own usually half-conscious opinions. It happens often that a pupil writes 90° by an angle which appears to be right-angled in the given figure and obtains a very simple solution which, unfortunately, is incorrect, because the angle in actual fact is perhaps 87° or thereabouts. And how many are there not who calculate $7 + 3 \cdot 5$ to 50 because they want to do things "in proper order"? Mathematics requires attention, not just with numbers and geometrical figures, but above all with one's own thinking.

It is necessary to have command of certain knowledge, e.g., that the multiplication $3 \cdot 5$ precedes the addition in the example above — only a convention, of course, but still important — or that the sum of the angles in a triangle is 180°. Some pupils who want to be on the safe side memorize a great deal of such things, but I usually emphasize that one can often help oneself with a simple example.

Those who memorize a bunch of things run the risk of mixing up the memorized facts and can be psychologically limited to the need to find the right thing in memory on different occasions.

A few examples: "What do I do here as the last step in the equation?" (The pupils point to the equation $16.3x = 0.5$.) "Should I divide 0.5 by 16.3 or the other way around?" — "What is the solution of the equation $5x = 20$?" — "x equals 4." — "Then you see what you should do in your equation; do the same thing there." — "Thanks, I know."

The simple example with the same structure as the more difficult problem has the pedagogical advantage that it is self-instructive — it contributes to the emancipation from authority.

Another example: "I forget, sir, is the area of a circle πr^2 or $2\pi r$?" — "What does r stand for?" — "Radius." — "What would 5r be, for example, a length or an area?" — "A length." — "And what would 6.28r be?" — "Thank you, I've got it now."

"Is the sum of angles in a triangle 180° or 360°?" — "Draw a right triangle." (The pupil makes a sketch.) "How big is the biggest angle?" — "90°" — "Is it 180° then?" (Hesitation.) "Double your triangle to make a rectangle; then I'm sure you can judge for yourself." (Figure 4.3.1).

I hold as very essential such practice in finding simple examples or simple ways back to basics. When pupils succeed with this, they gain a confidence which is invaluable. They may even go so far as to have confidence that they could reproduce, if necessary, the proof of an important theorem, without needing to use any special tricks of memory.

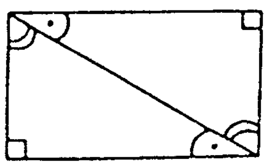


Figure 4.3.1

We then perhaps begin to achieve the most important goal of teaching mathematics: to have confidence in thinking itself.

5

MATHEMATICS AND SCIENCE

5.1 The Natural Sciences

The philosopher Immanuel Kant, and many others with him, claimed that "there is only as much true science in the natural sciences as there is mathematics in them."

Anyone who has had the experience of insightful knowing in mathematics, which was exemplified in the previous chapter, can easily understand Kant's statement. Long before Kant, Descartes emphasized that during his school studies he found the contents in only one subject to be certain beyond doubt: arithmetic and algebra.

Preparing for the study of physics, many have likely heard the advice to lay a good foundation in mathematics: "Then you won't have any problems in physics." It is both natural and a fact that mathematics has become a cornerstone of science. Yet one may ask himself whether it perhaps has not obtained too strong a position in school — at the cost of the study of nature itself — not least through the kind of tests which are given. In recent years even chemistry has become more and more "mathematical." Of course it is a question of balance. The question is if the balance is good today. An exaggerated mathematizing in physics and chemistry places obstacles to the natural studies: observation, the devising of experiments for systematic exploration, looking for answers and, if possible, finding a theory are cornerstones in natural science.

Let us compare mathematics and the natural sciences, beginning with the latter. As the name implies, natural science is a summary of knowledge of nature. Any understanding we might obtain must come first from observation of nature, observation which we do with our senses, possibly intensified with a microscope, telescope, oscilloscope, or other instruments. The observations collected must then be compared with each other and ordered. Certain observations seem to support each other; others seem to go against each other. In both cases thinking about the observations leads to new questions which guide the researcher to new experiments, new perceptions, new measurement; when research has come so far that an agreement, a common factor, a regularity has been found; then it is time for a theory which summarizes.

As a criterion for the value of a theory, we require that the theory be able to predict the outcome and results of new experiments; that the theory can be "confirmed" experimentally. When this has been done in a *large* number of cases, when it has been shown that the results are reproducible and not just the result of chance, the theory can expect to be recognized. Statistical methods must often be used to demonstrate that the obtained results are not the product of the researcher himself and his experimental arrangement but correspond to objective reality. In cases where results have not been able to be verified by other researchers, the theory gains only limited interest.. Natural science cannot be dependent on the practitioner: it is by its nature general. From Michael Faraday, considered by many to be the greatest experimental physicist throughout the ages, we have the following words worth thinking about, made in response to letters asking him to comment on the value of various new discoveries described by researchers:

> I was never able to make a fact my own without seeing it, and the descriptions of the best works altogether failed to convey to my mind such a knowledge of things as to allow myself to form a judgment upon them. It was so with new things. If Grove, or Wheatstone, or Gassiot, or any other told me a new fact and wanted my opinion, either of its value, or the cause, or the evidence it could give in any subject, I never could say anything until I had seen the fact. .
>
> (Faraday to Dr. Becker, Oct. 25, 1860, see L. Pearce Williams, *Michael Faraday*, London, 1965, p.27.)

Galileo himself, the father of modern physics, had, by and large, written a declaration of policy for the natural sciences when he wrote: Let us rely on demonstration, observation, and experiment. As early as in Galileo's works we find the role of mathematics pointed out: those who wish to solve scientific questions without the help of mathematics are

taking an impossible task. One must measure that which is measurable and make measurable that which is not.

Apart from whether natural science leads to knowledge in mathematical or other form, it is firmly anchored in a large number of observations. That the distance a ball rolls grows with the square of the time, as Galileo found, has been confirmed many times over by others than Galileo, and yet there is still an enormous difference between the formula $s = at^2/2$ for the rolling ball and a formula in mathematics such as the closely related $1 + 3 + 5 + \ldots = (2n - 1) = n^2$.

5.2 Mathematics

The formula $s = at^2/2$ has been arrived at empirically and means: under present laws of nature every ball rolling down a plane will cover a distance given by $s = at^2/2$. Almost all scientists consider scientific statements as statements of probability: the probability is currently very very low that a ball will show some other function for the distances covered.

The formula that the sum of odd numbers gives a perfect square is an entirely different matter. We have available a *proof* that the formula is true. The formula will always remain true. We are obliged with the physical "law" to time and again check its applicability — the mathematical formula is demonstrated once and for all. Anyone who has thought through the proof has it as an inner, conquered certainty.

It is of great pedagogical value to let pupils in a sixth grade class clip out angles in a paper triangle and see how big the sum of the angles comes to be. Thousands of people can do this and, measuring with protractors, obtain values around 180°. One could calculate averages and other statistical measures and come to the result 180.00° for the sum of angles. If mathematics were natural science, we would accept such a way of going about things. But it is clear to us that mathematical statement cannot be formulated as the result of experimental works. Intuition leads us to a way of proof, perhaps first of all to a supposed conclusion, but after that logical reasoning must confirm with certainty the correctness of the proposition. The contents of a statement are, of course, a function of the axioms which we have taken as a base.

Mathematics is a science which differs from natural science but which plays a large role in it and which has been and is a model to follow in many scientific contexts. It is relatively easy in a mathematical system to have the axioms which lie at the base of the system, but it was shown during this century that certainty is not absolute, even in so clearly delimited an area as the whole numbers. To obtain a sufficiently complete system for the whole numbers, such an encompassing system of axioms would be needed that (according to a proof 1931 by the German mathematician Kurt Gödel) one could make propositions, formulated in terms of the system's language, which can neither be proven nor disproven with the system's axioms themselves. Paul Finsler gave examples of these kinds of statements, where, nonetheless, our thinking can determine the correctness of the statement. In a certain sense, then, logical systems are relative. This might now be taken as a proof that mathematics cannot be based upon that certainty which logical thinking, according to our experience, leads to and therefore does not differ in character from other science. This, however, would be to underrate thinking. That mathematics cannot be "mechanized" into a logical system need not cast suspicion on thinking. On the contrary, it shows that thinking is the (only) archimedean point which mathematics has available. Gödel's conclusion demonstrates that mathematics is a conquest by our thinking. This does not stand in opposition to the entirely true statement that our sensory perceptions play a large role in our ability to pursue mathematics. They awaken concepts to awareness within us. We never really *see* a line, or a circle, not even a point, but any things stimulate us to arrive at these concepts and become conscious of them.

Here lies the fundamental difference between mathematics and natural science: mathematics springs forth from thinking itself; natural science must be based on observation. I do not, however, hereby consider myself to support Kant's statement that the laws of arithmetic are "a priori" truths. They are conquered by thinking in an alternating play with the senses.

5.3 The Two Parallelograms

During a lecture on natural science, Rudolf Steiner gave a good example of where the border line goes between mathematics and physics. I would like to relate it briefly here:

The principle of the velocity parallelogram and the principle of the force parallelogram seemingly have the same form: a particle which is given two simultaneous velocities as in Figure 5.3.1 obtains a velocity which in direction and size is determined by the vector sum of the given velocities (one of the diagonals in the parallelogram formed). Analogously: if a particle is acted upon by two forces as in Figure 5.3.2, then the resulting force is determined by the vector sum of the given forces.

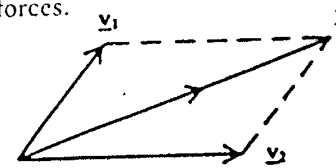


Figure 5.3.1

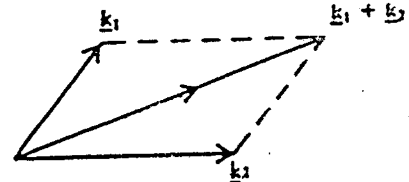


Figure 5.3.2

These principles come about in completely different ways: the resulting velocity can be derived from the concept of velocity and the given velocities, while the resultant force principle has been arrived at empirically. Where the force is concerned, it is *nature* which shows us that the parallelogram principle holds (at least so far...).

5.4 Descartes, Newton, and Gauss

Just as there are and have been thinkers who describe mathematics as a natural science, so are there researchers who have held physics, for example, to be a form of mathematics. The most prominent representative of this position was probably Descartes. The human spirit, according to him, brings forth the natural sciences the same way that it creates mathematics. Descartes presented his ideas on "universal mathematics" in the posthumously published work *Regulae ad directionem ingenii* (Rules for the guidance of the genius).

His method is basically to pursue physics according to the deductive methods of mathematics. Phenomena ought, according to Descartes, to be able to be derived from axioms, just as theorems in mathematics build upon axioms. The results of physics should be capable of being derived.

Here we must point out, in all fairness to Descartes, that he wanted to reduce physics to a theory on the motion of bodies. Through this he came to be one of the physicists who pioneered the way for the 19th century mechanistic world view: that natural science must be based in the end on mechanics (space, time and movement).

Descartes tried to practice what he preached. He claimed with determination, thereby directing himself primarily against his contemporary Pascal, that vacuum cannot possibly exist — except in Pascal's head. Pascal, who was certainly just as philosophically inclined as Descartes, let a number of experiments be carried out which confirmed the existence of vacuum and which led physics further along the right track concerning the question of vacuum. Descartes' deductive way of doing research fouled up here in a very noticeable way.

As further examples showing that the roots of mathematics lie in our own thinking, I would like to briefly mention

— that Pascal all on his own by the age of 12 had achieved a knowledge of geometry equivalent to a number of the theorems in Euclid's Elements;

— that Newton, at the age of 24 during the 1½ years when Cambridge University was closed due to the plague, conceived the major part of the pioneering mathematics which he later published;

— that Poncelet with no help and while a prisoner in Russia for over a year (during the Napolean war) laid an important foundation for the new projective geometry.

It is worth mentioning that Newton, in his famous work on the mathematical principles of natural philosophy, lays a system of definitions and axioms as a groundwork for his presentation. This axiom system has a different character, however, than a mathematical system of axioms. Newton describes among other things a number of research principles, for example:

> One must as far as possible account for similar effects with similar causes. For example, breathing in humans and in animals, the fall of a stone in Europe and in America, the light of the kitchen fire and from the sun, the reflection of light from the earth and from the planets.

And:

> in experimental physics one must hold true those propositions which have been won through induction from phenomena, until other phenomena come to be known...

We see that Newton clearly understood that the conclusions of physics are inductive, in contrast to the deductive results of mathematics.

I would also like to describe Gauss' measurement of the sum of angles in an enormous triangle whose points he placed on three mountain tops, namely the peaks of Hohenhagen, Brocken, and Inselberg, where the shortest side was 69 km. Gauss made the angle measurements with the aid of optical signals and found that the sum of angles was so close to 180° that the measurements could not possibly have shown any other value.

Did Gauss do this measurement in order to convince himself that the sum of angles in a triangle is 180°? Of course not. His question — in all certainty — must have been: when we measure an angle using optical instruments, for example, is it correct to use plane Euclidean geometry on the measurements? As we discovered in Section 3.10, in a spherical triangle the sum of angles is greater than 180°. Gauss' question concerned: which of the known geometries should we apply in a specific practical context? Only through experiment, Gauss rightly believed, could he get an answer to such a question.

Let us lastly return to the starting point for this chapter. It would be to the good of education in both mathematics and the natural sciences if the natural subjects did not become overgrown with mathematics. Because of the essential difference between research in the natural sciences and in mathematics, experiment ought to be given much room in physics and chemistry. The applications which currently are given so much attention in the natural sciences are often a numerical processing of previously given formulas. This does not invite creative mathematical training. I say this fully aware that problems in the natural sciences have been a superb source of inspiration, one of the very best, to mathematicians in pioneering new roads.

6

MATHEMATICS AS A SCHOOL SUBJECT

6.1 Alternating Between Practice and Orientation

Mathematics is a practice-subject, above all. If we limit ourselves to calculation, it is entirely a field of practice. On first going through an arithmetic formula or a theorem in geometry, pupils generally obtain only a first acquaintance, so to speak. They know the theorem no better than we know a person after a short first meeting. Even the first solved practical example in a new area gives for many only a hint of how the solution is actually done. Here is verified the old expression: "repetition is the mother of learning." Repetition in the form of practice, with as much independent exercise as possible.

We as teachers, on the other hand, may contribute to a first meeting with a mathematical topic by helping it to penetrate more deeply into the pupils than if we spoke like a book. Verbal teaching with its dialogue between class and teacher and between students themselves provides excellent opportunity for creating an atmosphere of excitement around a new element, so that an air of receptive readiness pervades the classroom. Such attentiveness means that impressions are stronger than otherwise. We forget more easily those things which we have placed somewhere without thought. If we want to help ourselves remember where we put a key, we should pay careful attention to and describe for ourselves what the surroundings look like where we place it, and imprint in our memory the picture of the key and the nearby surroundings.

By orienting a class on the "environment" surrounding a mathematical area, either by way of introduction or on a suitable later occasion, we give the pupils the opportunity to have living memories of knowledge. In Chapter 3 I have given a number of examples of such orientation: on a cultural epoch such as ancient Egypt for introducing arithmetic in different number bases, on the positions of leaves in plants for the discussion on Fibonacci numbers, and so on. Many problems can be made to come alive through glimpses from the biographies of the great mathematicians. Such orientation serves not only to strengthen the pupils' motivation; it leads also to the problem formulations which seem natural to the students.

It often seems natural to let historical presentations dominate in such orientation, but experience shows that historical descriptions seldom are particularly accessible in mathematics when they come right in the beginning. The pupils often have not yet made acquaintance with the mathematical material. An historical exposition generally has much greater effect when the pupils have become familiar with the new chapter's content and methods. In the case of number systems it is more effective to let the pupils themselves invent another system, different from the 10-system (with new symbols, see Section 3.1.1) than to directly present the ancient Egyptian system or the modern binary system. When introducing logarithms we made for ourselves a "table of logarithms," however primitive, with the aid of a curve and reading off the graph, and discovered some of the basic rules before the class was told about Jost Bürgi, a Swiss clockmaker who as early as 1588 constructed a shrewd logarithmic system, long before the Scotchman Napier published his work on logarithms in 1614. The classes are usually very interested in hearing about Bürgi's life and ideas and *can* listen actively, because by then they know roughly what a logarithm is and what difficulties come up during construction of a logarithmic system.

If one has something beautiful or exciting to show, which can awaken the children's astonishment, one should be careful not to bring it forth too early. "Astonishment should come last in the art of teaching," recommended Rudolf Steiner to teachers at a conference. Elements which are intended to make strong impressions on the pupils, which can surprise or fascinate them, must be prepared during classwork so that the pupils have the prerequisites for experiencing the intended item as a climax.

The ideal is probably an alternation between orientation in a subject and practical exercises in problem solving (including constructions in the case of geometry). From the orientation, problems spring up naturally and, if the problems are well chosen, give the class reason to ask questions around the problem-solving itself.

In geometry, orientation can quite nicely include elements from non-Euclidean geometries, so that pupils do not live with the same view as scientists did in Kant's time, that there is only one geometry, the Euclidean.

At university and schools of engineering one can meet students who in high school have gotten the impression that there are two kinds of series — arithmetic and geometric. They limited their knowledge of infinite series, one of the most fascinating chapters in the history of classical mathematics, to a dutiful calculation of the sums of a few series of arithmetic or geometrical type. They did not know that the concept of convergence of number sequences is the doorway to broad areas in mathematics.

An orientation on the creativity of some of the great mathematicians, Archimedes, Euler, Gauss, Pascal, Hilbert, and many others, ought certainly to be included in the mathematics course.

Concerning the importance of axioms, an introduction to Hilbert's contributions would have its rightful place. Hilbert held a series of lectures in the winter of 1898-99 on elementary geometry. He built up geometry from the ground floor, and it is interesting in an end-of-term class to compare Euclid's way of defining point, line, and plane with Hilbert's introduction. While Euclid, for example, describes a point as that which "lacks length and width," Hilbert introduces the basic elements point, line and plane directly in relation to each other through a number of so-called incidence-axioms (incidence = relationship). In Hilbert's work one can also find simple, instructive examples of how an axiom system can be tested for the three qualities which characterize an ideal system of axioms: the axioms should be complete when taken together, they should be independent of each other (so that one axiom cannot be derived from the others), and they should be non-contradictory.

Often in the literature, for example on Boolean algebra, one comes upon axiom systems which do not meet the requirement for independence, because the author preferred a more pedagogic presentation.

Proof that students may be interested in the foundations of mathematics, or in any case, that they can become interested, showed up clearly in the following episode which a teacher once witnessed in a sixth grade class. During a geometric lesson a couple of boys stood sharpening their pencils over the wastebasket to give them razor-sharp points. They stood unusually long but finally one said to the other, "You know, this is really hopeless. No matter what, we can never get a point!"

6.2 Quantity and Quality — Teaching on Different Levels

We live in a society in which quantities are given primary importance in many areas. We need only to think of the controversial subject "grade-point averages in the competitive school" to become aware of how society is organized such that quantities quite simply play a major role. In recent years quality has begun to be placed first; especially in the debate on our environment the expression "quality of life" has gained acceptance, even though often weak and though it surely invites the broadest interpretation.

Galileo's challenge to "measure that which is measurable and make objectively measurable that which is not" gained a following which he himself could hardly have guessed at. In various areas of society today there may be found a number of tests which make the claim to measure people's qualities and present the results in the form of easily interpreted point totals. Not long ago we could see examples in the newspaper of employment tests which showed how arbitrarily far the process of making the unmeasurable measurable has gone. In the 1800's there prevailed a widespread optimism within science, which by that time had already celebrated great triumphs — for example, the finding of the planet Neptune through a number of mathematical calculations. Qualities such as heat, color, and taste were thought to be amenable to "explanation" in terms of quantities in space and time. Even medicinal effects were thought to be capable of prediction by calculation along the lines of mechanics.

The quantity-minded stream of thought actually goes back to Descartes more than to Galileo. I cite from the introduction to a book on Descartes by Paul Valéry:

> Descartes is certainly one of those who bear responsibility for the life style of our times, in which everything is judged quantitatively. When the diagram was replaced by numbers, when all knowledge was put in the form of comparative measurements from which followed a de-valuing of anything which could not be expressed in arithmetical relationships, then something occurred which has had the greatest import in every area. To the one side is put everything measurable, to the other everything which cannot be measured.
>
> (From P. Valéry, *Les pages immortelles de Descartes*, Éd. Corréa, Paris 1941).

It would be carrying things too far to make Descartes scapegoat for the onesidedness in our culture. But it is apparent that he overestimated the value of arithmetic and that there is much truth in Valéry's words.

> A child who has learned arithmetic and done an addition according to the rules of arithmetic has in all certainty found in that doing all of what human thought is capable of finding,

claimed Descartes.

We can appreciate Descartes for his influence on natural science, which in its turn contributed to an impressive technology and through that to a strongly developed intellectual acuteness. Analytic geometry, introduced primarily by Descartes and used later in developments in the theory of functions, gave technology the prerequisites which were needed to apply the laws of causality to the construction of instruments, machines, and ships of all kinds.

But there has also existed and still exists another *scientific* direction apart from the quantitative, although not so patently successful in outward appearance as the mechanical-mathematical stream of thought. Atomic physicist Walter Heitler expressed in a lecture:

> One directs attention to qualitative phenomena... to qualities which have something to do with the observed object's wholeness. One of the most important of this philosophy's founders is Goethe, with his writings on natural philosophy.

Heitler gives "The Theory of Color" and "The Metamorphosis of Plants" as examples of Goethe's efforts in these directions.

> He directed his attention to the figure's unity and to the qualitative context. Goethe is the founder of modern comparative morphology within botany.

Why, you may ask, this long introduction before we get into the theme of teaching mathematics? Because I want with these historical examples to give a background to problems of which the mathematics teacher ought to be aware. What do we want from our teaching? Is it the main thing that students train up a given measure of tools and knowledge, i.e. do we place the quantitative aspect first? Schools of engineering want certain prerequisites, business school others, schools of medicine theirs, and so on. But should the high school or corresponding levels in our schools be preparatory for technical schools? If we want once again to place qualitative aspects and the individual's development during his and her time in school in the forefront, should teaching then be pressed into forms which are determined by point totals on exams?

Of course, all higher education demands a measure of preparatory knowledge, but it ought to be just as clear that every pupil should not need to go through the same preparation as candidates for higher education. The solution to the problem of how one can meet the needs of all the students must then be sought in some form of differentiation.

Should this differentiation mean that every pupil studies at his own pace and perhaps follows his own individual program, or can differentiation be best done within the framework of the school class? As far as I can see, the solution ought to have its emphasis in the latter alternative but with some elements of individual programs.

It ought to be possible to keep the class together and active around a basic framework of teaching and exercise materials.[1] But this requires that students be allowed to work at different levels of understanding and

[1]In the Waldorf School and other Swedish high schools the class is a fixed group which takes virtually all subjects together and which stays as a unit year after year until graduation.

achievement. In some of the teaching examples in Chapter 3 I have shown that there are problems which allow different levels of abstraction. They can be solved in an illustrative way, they can be treated more abstractly, or they can be approached in an elegant manner which directly leads to general results (see triangle numbers in 3.1, the rabbit problem in 3.3, the bottom sum problem in 3.5.2, the successive number sum in 3.5.5 and exercises in projective geometry in 3.10). A number of results can be achieved with very simple graphical methods *or* with possibly very demanding algebraic approaches (e.g. the problem of the meeting trains, the solution of simultaneous equations, etc.).

For spatial geometry an analogy to graphical solution can be that some pupils construct paper models and measure them, while others apply Pythagoreas' Theorem, solve an equation and so on. In such cases the pupils do not solve the problem on different levels as in the Fibonacci rabbit problem but rather in different ways. It is a matter of different choices of method. The advantage with problems of this type is also that groups which are working in different ways can become interested in each other's results. In this way a social element comes, without special effort, into the lessons.

But does not such choice of methods mean that groups work at different speeds? Of course it may happen, but then the groups can unite around the work which remains to be done. The teacher is also available the whole time and can help those who are having difficulty with methods which are more demanding than the illustrative or graphical.

This question of differentiation is often limited to mathematical problems of a quantitative nature, which concern calculating a length, an angle, an area, an intersection, etc. Would there not be reason to give considerably more time to *problems of a qualitative nature*?

What do such problems look like?

In principle they are problems in which investigation, not results in numeric form, plays the primary role — in which one needs to use a little fantasy, where the urge to build, construct, and shape can be put into action.

It is not difficult to find problems of this type in geometry, where quite naturally a whole category of problems involve constructions of different types. I need not here to further into this kind of problem but choose instead a few other types:

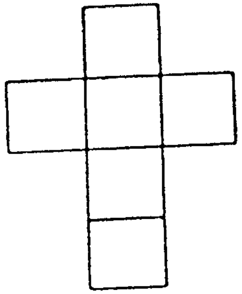


Figure 6.2.1

1. Exercise 1 in Section 3.7.3.

As we know, the network pattern for a cube may have the following shape (Figure 6.2.1): the four squares in a row can form the walls, the other two the top and bottom.

Problem: How many *different* networks can a cube have?

By two different networks we mean that the one network cannot be covered by the other even if the networks are cut out and turned around or turned over. For example, the following networks are the same (Figure 6.2.2):

Further, two adjacent squares in the network must have a common side, not just a common corner. A network such as in Figure 6.2.3 is not allowed.

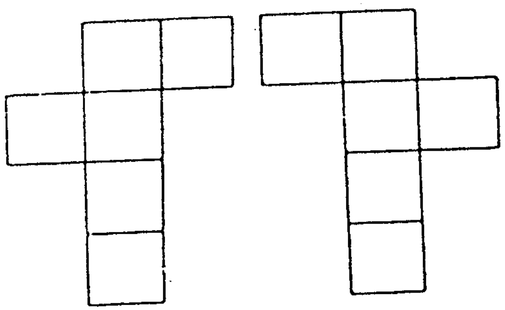


Figure 6.2.2

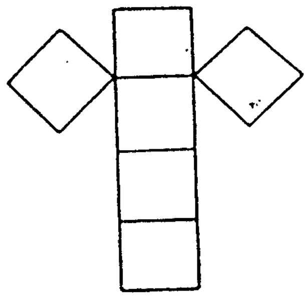


Figure 6.2.3

We have here a geometric-combinatoric problem of a constructive nature. It usually occupies the whole class, and quite intensively. Groups often form, and I emphasize that the problem concerns, above all, the question: how do we know when we have found *all* the networks which are to be found? How shall we determine that there are no other networks than those we have made? We come to the conclusion that some form of systematic ordering must be used in order for us to be able to determine how many networks the cube has. With this formulation it is not decisive how many networks any particular individual or group comes up with. The work is not meted out on a performance basis.

The person who finds "only" two networks may intellectually be the one who succeeds in finding a useful systematic ordering. (Concerning the number of networks, see Section 9.7).

This problem type can have many variations, even problems where the task is to find a single correct network for a solid. For example, we can ask: do both of the patterns in Figure 6.2.4 work for making a tetrahedron (three-sided pyramid)?

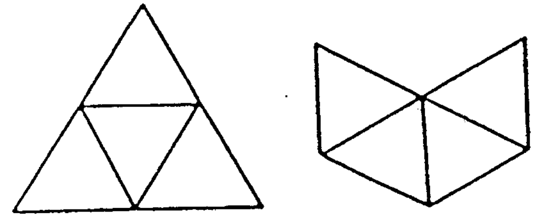


Figure 6.2.4

2. We wish to investigate how the diagonals of a quadrilateral can be determining for the quadrilateral shape. We take one of the following alternative conditions as a starting point:

(1) The diagonals are equal in length, at right angles, and bisect each other.
(2) The diagonals are equal in length and bisect each other.
(3) The diagonals are at right angles and bisect each other.
(4) The diagonals bisect each other.
(5) The diagonals are at right angles. One is bisected by the other.
(6) One diagonal is bisected by the other.
(The list could include still other alternatives.)

What kind of four-sided figures do we get? Can we summarize the results in a table which gives us an overview?

This problem might lead into a discussion on what are necessary and/or sufficient conditions concerning the diagonals such that we get a rectangle, for example. How is a romb characterized by its diagonals? And so on.

3. (From a period in spherical geometry.)

A person travels 1000 km south, then 1000 km east and finally 1000 km north. He is then back at his point of departure. Is there any starting point where this is possible, other than the North Pole?

This problem may seem to have the character of a trick-problem, but it gets the imagination into action (Solution in Section 9.9, exercise 5).

And finally I want to mention that projective geometry provides an excellent arena for problems of a qualitative nature. I limit myself to one example here and refer additionally to Section 3.10.

4. We begin with three corners A, B, and C, which form a triangle with sides AB, BC, and CA. We choose a fourth point, P, anywhere in the plane (but not on the sides of the triangle) and draw connecting lines to the three corners. We then get three points of intersection, X, Y, and Z with the lines AB, BC, and CA. What does the dual to this figure in the plane look like? (The dualization is based upon points and lines exchanging roles in the plane.) And what occurs when P happens to lie on one of the triangle sides?

It is much more difficult to find qualitative problems in arithmetic and algebra, since the material there is in fact numbers. The following examples may serve to illustrate:

1 a) Is it necessary for a whole number to end in 5 in order for it to be divisible by 5?
b) Is it sufficient for divisibility by 5 that a whole number ends in 0?

The students may motivate their answers with the aid of examples.

c) Is it necessary for a whole number to end in 0 in order for it to be divisible by 10?
d) Is it sufficient that the number ends in zero for it to be divisible by 10?

Please note: the purpose of these exercises is for the pupils to learn the concepts of necessary and sufficient conditions and to give them experience of what it means when conditions are both necessary and sufficient. We see often, in the most widespread contexts, necessary and sufficient conditions being mixed up with each other in everyday argumentation. What one person emphasizes as necessary conditions are understood by the other as a statement concerning sufficient conditions. For example, a person responsible for taking on a new history teacher says: "X is very knowledgeable in his field. He has even published books which have received good acclaim." The statement may be interpreted that the speaker considers X's knowledge of history as a sufficient merit for appointment. But what of X's teaching abilities? Isn't knowledge in the subject area only a necessary prerequisite, just as the ability to teach? And what would be sufficient conditions in this case?

2. The Pythagorean Theorem is one of geometry's most important theorems.

As we know it says that

$$c^2 = a^2 + b^2$$

where a and b are the sides of a right triangle and c is the hypotenuse. Is the condition "right triangle" here a necessary or sufficient condition for $c^2 = a^2 + b^2$ to hold? It is obviously a sufficient condition, as always in statements of the form "If..., then ...." Is "right triangle" also necessary? As a rule we overlook this question.

3. There exist, as some may know, so-called prime twins, i.e. pairs of prime numbers which are successive odd numbers: apart from the lowest we have

(11;13), (17;19), (29;31) etc.

(As far as I know, it is still not proven that the numbers of prime twins is infinite.)

Are there prime *triplets*, apart from (3; 5;7)? That is, can three successive odd numbers, larger than 3, all be prime?

A problem of this kind requires hardly any prerequisite knowledge at all and can immediately engage everyone in the class.

4. The well known problem of how a ferryman would get a wolf, a sheep and a head of cabbage across a river in his rowboat. (Or a fox, a chicken and a sack of grain.) He must assume that the wolf will eat up the sheep if left alone and that the sheep will eat up the cabbage. Further, he has room in his boat for only one thing at a time besides himself.

How does he get them across? This is a kind of combination problem, which can give the teacher good insight into the pupils' abilities, especially if they hand in their solutions in written form.

A simpler variation on the theme: A truck can carry 2000 kg. It is to move 5 heavy machines weighing 300 kg, 400, 500, 1200, and 1600 respectively. How should the transport be organized in the simplest way?

5. As qualitative problems I would also include combinatorics problems in which one needs to draw "decision trees" or similar diagrams, or where groups of letters are to be ordered alphabetically:

Which letter sequence comes just after and which sequence just before the sequence

BCDAE

when all possible sequences made up of the five letters A-E with each letter appearing only once are arranged alphabetically?

The majority of students have an easy time finding the next following letter sequence in such examples but rather more difficulty in finding the preceding sequence. Exercises in going *backwards* in letter groupings, number sequences etc. require more effort of will in one's thinking.

6.3 Should All Pupils Take Mathematics

Based on experience in the Waldorf schools, my recommendation would be that all pupils be given the opportunity to take mathematics during the whole of their school education, which in the Waldorf schools is twelve years. But the answer to the question in the heading is naturally dependent on the kind of school, the organization of the courses, possible integration of school subjects, and above all on the goals of the school form. A school which seeks to give pupils a basis for specific kinds of vocational training differs essentially from a school which seeks to give general preparation, an education apart from the pupil's later education or choice of work. The Waldorf schools belong to the latter type of school. Their primary goal is to develop insofar as possible thinking, feeling, and will into an inner harmony so that youth can go into life with ability to observe and listen to the surrounding world, weigh pros and cons, form mature judgements and make well-grounded decisions. The Swedish elementary schools have the same goal.

In my opinion mathematics has so many opportunities for giving pupils valuable inner abilities that place for it ought to be found in the

schedule during all the years in school, even if possible in schools which prepare for specialized work. For safety's sake, I would like to emphasize that I do not consider mathematics to be more important than other school subjects. Every activity has its special value. Any one of the subjects in school may be the field of activity which at a given moment is the very most important for a particular pupil.

Mathematics contributes with qualities which cannot be replaced by exercises in other subjects. The particularly valuable aspect of mathematics is its opportunity, properly pursued, to develop confidence in one's own thinking, a confidence which is built up through experiences of inner certainty of knowing.

Mathematics is, of course, not by any means the only area where one can develop powers of thought, but it would be going too far in the other direction to consider courses in foreign language as an equivalent to training in mathematics. In translating from one language to another, even to or from Latin, considerably more elements of "convention" come into play than in solving problems in mathematics. Those who would draw a parallel between the rules of mathematics and grammar should consider that grammatical rules often have exceptions, which clearly shows that a language does not build upon logic in the same manner as does mathematics. The exceptions, in fact, are that bring language alive, and the fasciating aspect of language studies is perhaps trying to speak and write so that it sounds truly genuine.

That there are pupils who are gifted in language but have great difficulty in both arithmetic and geometry even drawing the figures — confirms that language and mathematics direct themselves to youngsters in different ways. In a class which I was fortunate to know especially well, the best student in French could not, even after help, see depth in a two-dimensional drawing of a three-dimensional solid (Figure 6.3.1).

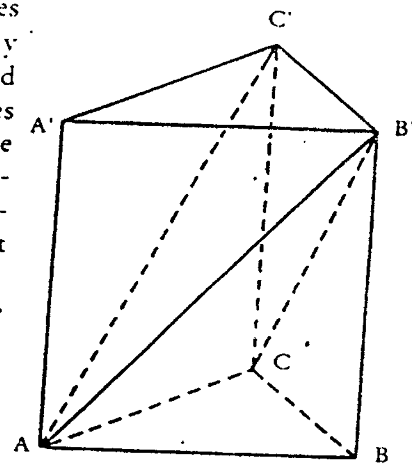


Figure 6.3.1
Can you see the three-sided prism with triangle ABC as the bottom and A'B'C' as the top? Can you also see three tetrahedrons which this prism is divided into by the three diagonals AB', AC', CB'?

In other individuals the allocation of talents is just the opposite.

What we shall now take up is the question of how mathematics instruction can be organized so that it becomes, as far as possible, rewarding for *all* pupils.

For this, the beginning, the first year of school, should be given great importance. It is in the first year of school that children are asked to use their number concepts systematically and where they meet the four basic operations of arithmetic. From colleagues in the lower grades I have understood that even the natural numbers in sequence work to bring order to the inner life of the child. The child meets something objective and great in the world of numbers. Here as in the drawing of shapes and forms, the basis of geometry, an inner certainty begins slowly to sprout, completely unconsciously, of what later will become a conscious asset.

As we know, proportionality is a very important area in everyday mathematics. Its applications are so common that familiarity with proportions should be included in every child's education. Weighing something over, choosing between alternatives is, in fact, seeing the proportions on a qualitative field between action and effect.

When we now take up the question of how mathematics instruction can reach and work on *all* pupils, we usually think primarily of those youngsters who lack aptitude or interest or both, and who experience the daily lessons as a repetitive failure, regardless of how competent the teacher may appear to be. What can we do for these so-called underachieving pupils? Shall they suffer through a subject, perhaps during their whole school lives, because the subject has much to offer other pupils who manage to "keep up" or even to be "clever"?

This question is enormously important. A school subject must naturally not be allowed to be a cause of suffering! Nor to be something dreary or indifferent.

It is a fact that usually more than one pupil in an undifferentiated class, perhaps as many as 5 or 6, have substantial difficulty with mathematics and with logical operations (it usually shows up even in grammar exercises). We know of two measures for solving this:

(1) to give the pupils in question very simple examples in addition to the basic course, a course which hopefully they will get something out of together with their classmates — nota bene: assuming the basic course can be pursued on different levels of abstraction (see e.g. Sec. 3.3.2);

(2) that these pupils get help from a special teacher as early on as possible.

In most cases both measures are needed. It is difficult to predict how successful the results will be for the individual pupil.

I have had pupils who had great difficulties when I received them in the beginning of the eighth grade but who went slowly and steadily forward. During the twelfth grade, the last year, a pupil awakened and developed his talents (still within the framework of easy problems) unexpectedly well. Such pupils were not as untalented in the subject as one might have thought; they were dreamers and first awakened to consciousness in their thinking at the end of their schooling. One of these pupils, as I heard from his elementary class teacher, had as a newcomer been looked for by his parents a few hours after the end of the first day of school. He came home four hours late. What had he been doing? Standing and watching an excavating machine digging, without noticing the time.

Other pupils have not awakened nor come to the kind of involvement in the subject which one had hoped for. A failure? Perhaps. Without it being taken as a kind of general excuse I would like to mention here that it has happened in *rare* cases that the pupil, several years after leaving the school, has succeeded in working up an ability in mathematics or in some other area where logic plays an important role. One pupil who had difficulty getting better than just satisfactory results on his exams, studied several years later at university and got his degree in mathematics with good results.

The awakening of thinking *can* come suddenly. A Norwegian teacher told me once about a boy in the fourth grade who was still very weak in arithmetic and drawing geometrical forms. He had, on the other hand, a rich fantasy, so flowing that its effects often spread out over the whole class and caused the teacher problems. One morning during the spring of fourth grade the boy came to the teacher and said: "Teacher, I can do arithmetic now!" The teacher could not help being doubtful, but it turned out that after that day the boy worked practically every problem correctly and became the best in class. He later took the highest academic honors and is now active as a professor of mathematics!

Perhaps the most famous example of success after leaving school for a pupil with low grades is Einstein, who is said to have "failed" in arithmetic. These sunshine examples are not intended to cover up the difficult problem we face, but they have their place in showing that we need not resign ourselves if the efforts to help weak pupils seem to give little result. We need not have the expectation that they will become professors of mathematics, but we can sharpen our powers of observation in order to better notice the progress.

A number of experiences seem to show that the greater variation we can achieve in our mathematics problems, and the more we can pose problems which do not require very much previous knowledge, the greater are the chances that weak achievers will be stimulated to learn important basic examples.

The need for concreteness in education seems to increase over the years. Some pupils in the sixth through eighth grades color geometric figures in order to get a clearer grasp of the concept of area. Areas can also be compared by cutting a shape into pieces and putting it together again in a new form.

And how many pupils mix up πr^2 (the area of a circle) and $2\pi r$ (its circumference — despite the mathematics teacher's efforts with the dimensional comparison of r^2 and r respectively! Not until in metalwork, where the task is to make a bracelet with copper sheet as the material, do the work and the calculation problem become concrete enough for some students so that the formula for the circle's circumference takes on reality.

There is no doubt that desk calculators have a part to play in high school, but in elementary school I think that the advantages of the calculator in certain parts of the curriculum do not outweigh the disadvantages which follow in the calculator's wake: above all, a growing dependence on tools whenever calculations have to be made. I am convinced that a sound ability to carry out numerical calculations with pen and paper is a necessary prerequisite for achieving the capability and assuredness which is desired in mathematics.

The majority of pupils in the eighth and ninth grades at Kristoffer School have not shown unwillingness to do their own numerical calculations. Many pupils in these grades have the need to recapitulate and practice elementary procedures in arithmetic. I fear that the pocket calculator would have covered up this weakness, and likely made it worse, if it had come into general use.

7

BEING A MATH TEACHER AMONG TEENAGERS

7.1 Phases of Development before Puberty

There are three prerequisites for successful teaching. The teacher must:

1. be familiar with developmental psychology and knowledge of man for the stage the pupils are at;

2. know the subject;

3. get to know the pupils in the class as soon as possible, not only in the classroom but outside as well.

Here are some reflections on the first point, primarily with respect to teaching mathematics.

The child goes through a number of thresholds during his development. Not the first is maturity for school which in general comes along with the loss of milk teeth about the age of 7. Approximately in the third grade, about the ages of 9-10, the child wakes up considerably to his surroundings after earlier having been woven into them. But first at the age of 12-13 does the ability to form independent judgements begin to awaken; the child then also has significantly expanded intellectual resources. Prior to sixth grade most pupils have no interest in mathematical proofs: the inner resources for this are quite simply in a latent state.

Children in the lower (1-3) and middle (4-6) grades want to come into concrete contact with the problems which are to be worked. Experience from practical participation plays a decisive role for them. Piaget's research confirms in large measure the basis for the curriculum which Rudolf Steiner outlined for the first Waldorf school in Stuttgart in 1919. Piaget demonstrated that the child lives "in the stage of concrete operations" up to at least the age 12 years, possibly a few more years. First from about 15 years of age is there a maturity for tasks within the realm of "the stage of formal operations."

Piaget has described a number of examples which show how much in vain it would be to begin with abstract things before children have the maturity which is required.

In the Waldorf schools physics is taught beginning in the sixth grade, chemistry from the seventh. The pupils are trained methodically to follow attentively along in what is happening in simple experiments and afterwards to describe them. Eventually some of the experiments can be contrasted with each other and give rise to thoughtful reflection, so that the class learns to compare and, in time, to draw simple conclusions. The important thing is for the work to be based upon experience of a process. The class is asked to draw the experimental apparatus and to try to describe "what happened" as simply as possible.

This drawing and describing (first verbally, then written) comprises an important preparatory stage for a more concept-oriented penetration of natural phenomena.

In a corresponding way teaching in mathematics should go from the concrete experience to concept and context. The rule "from hand to heart to head" makes possible meeting the children on their own level. That which the hand does — draw, cut out, form, shape or build — gives the child inner experience. In mathematics these can include many impressions of beauty in symmetries or other relations. The activity has then entered into the child's emotional life, gone from hand to heart. Experience shows the best, most fruitful questions for concepts and explanations come from pupils who do not ask out of a quick intellectuality but out of a need to go on from their emotional involvement to a clarity in thought.

In an article in *Scientific American* (No. 11/1953) "How Children Form Mathematical Concepts," Jean Piaget writes:

> A child's order of development in geometry seems to reverse the order of historical discovery. Scientific geometry began with the Euclidean system (concerned with figures, angles and so on), developed in the 17th century the so-called projective geometry (dealing with problems of perspective), and finally came in the 19th century to topology (describing spatial relationships in a general qualitative way — for instance, the distinction between open and closed structures, interiority and exteriority, proximity and separation). A child begins with the last: his first geometrical discoveries are topological. At the age of three he readily distinguishes between open and closed figures: if you ask him to copy a square or a triangle, he draws a closed circle; he draws a cross with two separate lines. If you show him a drawing of a large circle with a small circle inside, he is quite capable of reproducing this relationship, and he can also draw a small circle outside or attached to the edge of the large one. All this he can do before he can draw a rectangle . . . Not until a considerable time after he has mastered topological relationships does he begin to develop his notions of Euclidean and projective geometry. Then he build those simultaneously.

I permit myself to doubt strongly the last statement. Need and ability in Euclidean geometry, according to my experience, come before abilities in projective geometry. But as a whole the quotation points clearly to the importance of a didactic road from the child's personal field of experience to concept formation via the inner experience.

7.2 Puberty

When the children near puberty and enter into it, their repertory of feelings is considerably broadened. Interest knows no bounds, and it is important to direct it out into the world, away from the ego. Even the intellectual powers increase strongly in many pupils. It is striking how willingly 15-year-olds like to get into discussions with teachers. There are pupils who always get the last word in. The will to discuss sometimes gives an impression of an instinct to sharpen the intellectual tools.

In this phase of life, of liberation from the adult world, of confrontation with the problems of the time with questions of life's meaning, and while filled with sensitivity, uncertainty, aggressiveness but also idealism, it is distasteful for many pupils to "plod" through technical concepts of arithmetic, to recapitulate and firm up so-called basic skills. Some become so out-of-tune by such routine exercise that they can't even bring themselves to get started during a lesson period. On the whole, youth in this age want to test their powers on new problems and use their cleverness in a *conscious* way.

Recapitulation and practice must be done in such a way that one simultaneously brings in something new. The young child's unconscious demand for concreteness in work has been transformed to a demand that the content of instruction be motivated, anchored in reality, not necessarily material reality. Youth have a right to get motivation from adults. They want in fact precisely to train up their ability to formulate motives for their own actions. The more the theme of classwork can come alive within the youngsters' own thinking without the teacher having to make introductions and expositions during the work, the better.

For some pupils, ability in mathematics seems to take a step backward during puberty. Self-confidence leaves them during classwork, just as it probably fails them during their free time. It is enough for a problem to "look hard" for some form of resignation to set in. During some lessons it is more important to find the right psychology than to speak to the subject. Mathematical activity requires, as we know, both time and patience, and it must be truly difficult for a pupil to bring himself to solve a problem if his self-confidence fails him in the classroom. A humorous word from somewhere can break the spell, or perhaps a pause with a few folksongs? Such a pause shows in any case that the teacher is not so incredibly intent on "making use" of every minute. Classes usually notice the teacher's degree of seriousness in different situations. Many can also separate teacher and subject, but not all. Some like a subject but do not have much sympathy for the teacher, for others it can be the other way around. A 16-year-old who had difficulty in mathematics said of his teacher, "NN is all right, but he has the subject against him."

Perhaps mathematics teaching has an important task just in this phase of puberty, when the pupils often have to fight so hard to reach objectivity. If one can relate stories from the pupils' own early years in school, or from other childrens' first years, it usually gets a good hearing: teen-agers can recognize very much of themselves in a problem situation which *they* have on another level than the young child, without feeling pointed out.

How wishful thinking can lead to mistakes comes forth quite nicely in the story of an episode which occurred during the registration of a girl who was to begin first grade the coming spring. She and her mother sat together with the teacher-to-be. After a while the teacher wished to feel out the girl's abilities in arithmetic a little bit. He said: "Five soldiers stand guard by a road. How many guards are left if four go away?"

The girl thought a minute, then answered "two." The mother paled and looked worried. The teacher could not have considered the answer as any great failure, but the mother later asked her daughter, how could she say "two"? "Why, mama, you know I felt so sorry that the soldier would have to stand all alone that I said two."

The emotions weigh heavily in the 14-16 year-old, making it difficult for many students to *want* to think about a problem. They would like to recognize the problem and be able to solve it without effort using some method they already know. Yet it is just precisely the unusual problems which would jolt the pupils out of their routine and infuse the power of will into their thinking. A few motivational words from the "grassroots' level" on why we take the time to solve a particular type of problem usually fall on good earth, because youth have a sense for any training which concerns them existentially, which might mean something for them even after leaving school.

7.3 Genetic Teaching

There are schools where so-called genetic teaching has been widely practiced. I am thinking of Martin Wagenschein, who in a number of publications emphasizes the importance, even the necessity, of a "genetic" teaching where pupils — sometimes with a little help from the teacher if needed — out of their own activity solve problems from the start by building up experience of simple, but in the given context, appropriate examples (see earlier sections on the importance of examples: 3.5.1-3.5.4 and Chapter 4). Wagenschein describes a project on the question, "Are there infinitely many prime numbers?" Thirteen boys and girls from different countries, between the ages of 14 and 17, who studied at a Swiss "free" school were asked, without special prerequisites or preparation, to take on the prime number problem which Euclid so elegantly solved. Is the number of prime numbers infinite? The group used five 60-minute periods to accomplish the task. Their logbook looked briefly like this:

The first lesson was needed for the group to become completely familiar with the question.

The second lesson was used for a discussion and investigation as to whether
a) 2n + 1 is a prime number generator
b) 6n + 1 or 6n - 1 always gives a prime number
The second question was broadened by turning around: Does a prime number always have the form 6n + 1 or 6n - 1?

The third hour: Discussion of the concepts necessary and sufficient conditions was absolutely necessary! Starting point in simple examples such as: Are all inhabitants of Switzerland Bernese (i.e. from Bern)? Are all inhabitants of Bern Swiss?

Continued discussion of the problems connected with 6n ± 1, it was a matter of making sure that all in the group fully understood the progress that had been made.

Fourth meeting: The pupils were asked to write down all the results they had come up with so far. One girl was practically at the goal but did not succeed in formulating the analysis of two cases which form the key point in Euclid's proof; at least she did not succeed in making her own insight understandable to the others.

The fifth hour was wholly directed toward the final formulation. The girl in question later wrote a letter in which she said, among other things: "When we after several days had solved the problem, we were so proud, as if the prime number problem had tormented us our whole lives and we were the first to find the proof."

Pride ought certainly not be the goal in itself, but if we look at the essence of this citation we see that it expresses satisfaction over that which one has achieved through his own work and effort.

To use an analogy: there is a difference between reaching the top of a mountain by car and getting there on foot.

The feeling of the joy of working is not easily found for a pupil with difficulties in the subject. If the evaluation of work in mathematics is primarily based upon written tests, then some pupils can feel pre-destined to getting less encouraging evaluations. The method used throughout the Waldorf schools where each pupil keeps his own workbook can be of great help: for some in the class it is considerably easier to sit in peace and quiet at home and think through course material, and they can present it in an individual way, even in mathematics, where the opportunity for personal ideas in the subject must be limited in comparison with orientation subjects. As pointed out in Section 6.1, the opportunities for individual contribution are greater to the same extent that the mathematics lesson includes elements of orientation.

The keeping of workbooks as a rule gives a kind of satisfaction over seeing a finished job, which stimulates continued efforts. But what school can afford to give five whole hours to the solution of a problem like Euclid's theorem on prime numbers, when in addition the result leads to no practical application? In what kind of school can genetic teaching find its rightful place, regardless of how appealing it may seem?

Perhaps it can be included, despite the above, to some extent; or more than is found in our schools today. The time is well spent if it is used as suggested by Wagenschein's examples. The student group familiarized themselves with the problem, put down some simple "theories" in the beginning and were motivated to test them out. After a number of tries down paths which were not successful, they eventually found the right track. It does not matter that "tips" or "hints" from the teacher may be necessary. During the actual seeking the group felt the need to study the concepts of "necessary" and "sufficient" conditions, concepts which are of great importance in mathematics.

Finally, the group did the work of formulating the proof. Doesn't such work often give significantly more for the time than time spent on routine tasks? It ought to be possible to select a few suitable course elements for such a genetic study.

Another counter argument worth considering is: will not genetic education activate primarily those pupils who are already bright? Perhaps some of the pupils will end up being more or less spectators? This risk exists, and it is up to the teacher to contribute to stimulating everyone. Genetic lessons often require more preparation than other forms of work, for the teacher as well as for the class (not least socially). The brightest pupils must develop a sensitivity for the social, more than in normal forms of work, so that they don't speak out too early, unasked by the class, and take away their classmates' joy of discovery on their own. When social relations begin running smoothly, the group work can be organized so that some pupils become "assistants" to others. Those who are the helpers will certainly come to experience that they themselves understand the results better when they are in the position of answering questions from their friends and giving them helpful hints. Even the act of understanding what a question means can be worthwhile exercise.

7.4 On Kinds of Abilities

I have already mentioned (Section 6.3) the occurrence of different kinds of talents for languages and mathematics. Included in point 3 above — knowing the individual pupil — is developing as quickly as possible a picture of his or her prerequisites for the subject. Just as there are pupils with an aptitude for language *or* mathematics, it sometimes occurs that a pupil has a talent for arithmetic *or* geometry. Most mathematically talented pupils find both of these branches easy, but in individual cases a pupil with good aptitude for geometry can have difficulties in work with numbers. This indicates that arithmetic and geometry direct themselves to different fields of ability, just as, for example, language and geometry.

In an eighth grade class one of the pupils was called "professor." This honorable title had been given to him after a number of prominent contributions in arithmetic calculation, and I was naturally interested to follow his further development. It turned out the next year in ninth grade, where the problems more so than earlier appealed to individual initiative, that the "professor" quite often asked for "hints." His ability had mainly been of a reproducing, receptive kind and no great source of heuristic, inductive thinking was to spring up during his remaining

school years. In the higher grades this pupil, for better or worse, went in for memorizing methods of solution.

Other pupils who had not shown themselves active in the lower grades, succeeded in coming more into their own in their later school years. Through suitable choice of problems the teacher can entice ideas out of pupils who usually do not express themselves so much. Faced with unfamiliar, perhaps surprising problems, "clever" pupils often look for a rule or other knowledge in memory, while less advanced youth tend more to use their "common sense" to find an opening for the problem solution.

8

ON THE CURRICULUM AND THE GOAL

The school seeks to develop capabilities of spirit and soul in the pupils so that they can feel an individual responsibility when they are faced with tasks in society. This feeling of responsibility ought to be the inner strength with which we take in a new task, see the prerequisites for its completion, and come up with creative ideas on how the work can best be carried out. Understanding as quickly as possible the problems involved in a task requires thought and insight. All school subjects should contribute to the development of the pupils' power of judgment. In mathematics particularly there are opportunities for practicing on clearly posed problems and on problems of the most varied degree of difficulty. One can emphasize the analytical aspect by putting a problem within a long text, or one may emphasize the problem-solving aspect which puts demands on constructive thinking. Opportunities for differentiating abound and we ought to be able to stimulate each pupil to work.

Should, then, knowledge in a subject such as mathematics be considered as a by-product to the development of inner abilities? No. To be sure we wish to see the child's inner development as the primary goal, but we would not get very far if we did not give the pupils opportunity to achieve knowledge. Acquiring knowledge is in itself an important exercise and a prerequisite for solving complex problems. Learning things is one step in the pupil's development as an individual. And we can be assured that every child comes to school with a strong, though largely unconscious, desire to learn things which have meaning and importance for later life.

The curriculum should be an aid to the realization of the goals. In mathematics the emphasis must be put on problem-solving because mathematics means independent thinking and this requires training. As

emphasized in Section 6.1, however, orientation on cultural epochs, on the lives of mathematicians, and on methods has its important place as a source of inspiration for the pupils and the teacher. Pupils get important historical and human perspectives. They can see how mathematicians solved problems and how they developed new methods.

Orientation itself is thereby a kind of practice and can become a stimulating introduction to more routine problem-solving. It is important that pupils get the chance to think through such orientational concepts and methods and to summarize what they have learned, preferably with their own reflections. Experience alternating orientation with problem-solving has been good.

It has often been said that mathematics curriculum should give greater space to modern areas and not take its material exclusively from times before the nineteenth century. The choice of orientational and practical training elements must be made with great care, and Pólya is here the best example. It is a question of using time wisely and simultaneously hitting on that which is pedagogically effective. A problem from ancient Greece can be equally fruitful as one from our times. Naturally there should be room for orientation on new advances in mathematics, but the overriding criteria should be that the material encourages development in the pupil.

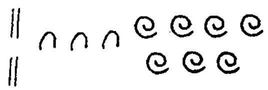
If one requires that the last years of school be dictated by the demands of university and college, then examinations and scores will come to restrict the bounds for general developmental elements in the curriculum. Perhaps it would be possible to arrange on a wider scale preparatory courses in the respective subjects at universities for those school children who wish to study further. Through this the pressure on the schools would diminish to the advantage of the majority of students who do not intend to follow the academic path. In his book *Matematik för vår tid* (*Mathematics for Our Times*) Professor Lennart Carleson points out that even within modern areas of mathematics one ought to be able to "give problems which ... measure mathematical abilities in a more genuine sense than tests of traditional type."

9

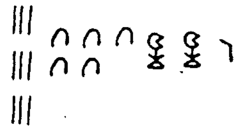
ANSWERS AND EXPLANATIONS TO THE EXERCISES

9.1 Exercises in Section 3.1.8

1. a) 734 and 12059 written in Egyptian hieroglyphic:



and



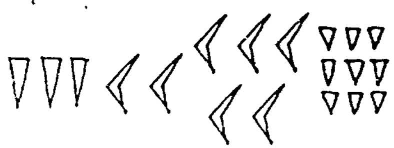
b) written in cuneiform:

$$734 = 1 \cdot 600 + 2 \cdot 60 + 1 \cdot 10 + 4 \cdot 1$$

which is represented

$$12059 = 3 \cdot 3600 + \text{a remainder of } 1259$$
$$1259 = 2 \cdot 600 + \text{a remainder of } 59$$
$$59 = 5 \cdot 10 + 9$$

which is represented



2. a) $39 = 124_5\ (1 \cdot 25 + 2 \cdot 5 + 4 \cdot 1)$
b) $150 = 1100_5$

c) $795 = 1 \cdot 625 + \text{remainder } 170$
$170 = 1 \cdot 125 + \text{remainder } 45$
$45 = 1 \cdot 25 + \text{remainder } 20$
$20 = 4 \cdot 5 + 0$

from which we get $795 = 11140_5$

3. Yes, every natural number can be uniquely represented in the base-5-system. For example, to count an unknown number of buttons with the help of the base-5-system we arrange the buttons in piles of five. If doing this gives 3 buttons left over, then the last digit of the number must be 3. If no buttons are left, this digit must be zero. If the number of buttons is less than 25, the 5-position is determined by the number of piles. If there are 25 buttons or more, we arrange the fives-piles into groups, each containing 5 five-piles. The number of five-piles left over from the groups then gives the 5-position digit. And so forth. This representation is unique.

4. a) $34_5 = 19$ b) $230_5 = 65$ c) $304_5 = 79$
d) $10110_2 = 22\ (16 + 4 + 2)$ e) $11011_2 = 27$

5. The 5-system multiplication table:

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	11	13
3	0	3	11	14	22
4	0	4	13	22	31

6. a) $114_5\ (16 + 18 = 34)$ b) 111_2 c) $1111_2\ (21 - 6 = 15)$
d) $1111_2\ (5 \cdot 3 = 15)$

7. The sum is $10;010_2 = 2 + \frac{1}{4} = 2.25$

8. Each new chord is to be drawn so that it cuts all previous chords in new intersections (otherwise the number of areas will not be maximal).

Imagine that we begin to draw the new chord: each time it comes to an earlier chord we have obtained a new area, since an existing area is divided into two parts. When we arrive finally at the circle's edge we get the last new area.

The number of new areas is thus = the number of old chords + 1
= the new number of chords.

No chords	gives	1 area
1 chord	gives	1 + 1 areas
2 chords	give	1 + 1 + 2 areas
3 chords	give	1 + 1 + 2 + 3 areas, etc.
n chords	give	$1 + 1 + 2 + 3 + \ldots + n = 1 + \frac{n(n+1)}{2}$ areas.

9. Beginning with the second, each figure can be divided up into a square number plus a triangle number. If $k_1, k_2, k_3, \ldots$ and $t_1, t_2, t_3, \ldots$ are the successive square and triangle numbers respectively, then we get the following simple expressions for the pentagonal numbers $f_1, f_2, f_3, \ldots$

$$f_1 = 1^2$$

$$f_2 = k_2 + t_1 = 2^2 + \frac{1 \cdot 2}{2}$$

$$f_3 = k_3 + t_2 = 3^2 + \frac{2 \cdot 3}{2}, \text{ etc.}$$

The general formula, according to Section 3.1.5 then becomes

$$f_n = k_n + t_{n-1} \qquad \text{for} \quad n = 1, 2, 3, \ldots \qquad (\text{with } t_0 = 0)$$

$$f_n = n^2 + \frac{(n-1)\,n}{2} \qquad \text{or} \qquad f_n = \frac{3n^2 - n}{2}, \quad n = 1, 2, 3, \ldots$$

from which

10. All the numbers generated by the polynomial $x^2 + x + 41$ ($x = 0, 1, 2, \ldots$) land in the corners of the spiral square. This is because the number of steps from one corner to the next following corner forms the sequence 2, 4, 6, 8, ..., which is identical to the increase in the polynomial when x increases by 1:

$$(x+1)^2 + (x+1) + 41 - (x^2 + x + 41) = 2x + 2 = 2(x+1)$$

from which it follows that the increase is 2, 4, 6, 8, ... for x = 0, 1, 2, ...

9.2 Answers and Explanations to the Exercises in Section 3.2.8

1. The ordering principle for the 10 paths in Figure 3.2.2 is:

"Take an avenue as early as possible and proceed along it as far as possible." (The avenues run downward to the right, see Figure 3.2.1 and corresponding text).

If one wishes, the different paths can be characterized by letter codes. To that end we let an "a" stand for one block length along an avenue (anywhere) and a "b" for the same anywhere along a boulevard.

Path no. 1 can then be denoted by the code aaabb. (One goes first 3 blocks along an avenue and then 2 blocks along a boulevard.)

Path no. 2 is aabab.

One can in fact, *without looking at Figure 3.2.2.*, note down all possible paths by letting the letter combination

aaabb

take on all nine of its successors in alphabetical order. (There are to be 3 a's and 2 b's in each combination.)

The sequence then becomes:

1. aaabb	6. abbaa
2. aabab	7. baaab
3. aabba	8. baaba
4. abaab	9. babaa
5. ababa	10. bbaaa

2. According to Section 3.2.4 we have

$$(1+x)^n = \binom{n}{0}1^n + \binom{n}{1}1^{n-1}x + \binom{n}{2}1^{n-2}x^2 + \ldots + \binom{n}{n}x^n$$

Since all powers of 1 are 1 we can simply write

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \ldots + \binom{n}{n}x^n$$

$x = 1$ now gives $$2^n = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n}$$

The sums of the binomial coefficients in the rows of Pascal's triangle thus form the sequence 1, 2, 4, 8, 16, etc.

3. Seven people can be seated in $7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040$ different ways along the side of a table.

4. $\binom{7}{4} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{4 \cdot 3 \cdot 2 \cdot 1} = 35$ committee combinations.

5. The formula $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$

gives $$\binom{n}{n-k} = \frac{n!}{(n-k)!\,[n-(n-k)]!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{k}$$

6. Number of chords = number of connecting lines between pairs of points
= number of ways to choose pairs
= number of ways to choose 2 points out of 10 points

$$= \frac{10 \cdot 9}{2 \cdot 1} = 45$$

7. Number of triangles $\binom{12}{3} = \frac{12 \cdot 11 \cdot 10}{3 \cdot 2 \cdot 1} = 220$

8. a) The number of letter groupings (permutations) is

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$$

b) The number of permutations with repetition allowed is

$$5^5 = 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 = 3125$$

9. We call the 200 tries the thief makes a trial series. We know from Section 3.2.7 that each try can be done in 1024 ways, of which 1023 will not succeed in opening the lock. 200 tries made randomly each time (so that any given try combination may be repeated one or more times) leads to

$$1024^{200} \text{ possible tries}$$

among which 1023^{200} are unsuccessful trial series.

The possibility that the thief is unsuccessful in 200 attempts is thus

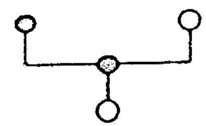
$$s = \left(\frac{1023}{1024}\right)^{200}$$

The desired risk is the probability that the thief succeeds, which is the complementary possibility of s, i.e. 1-s. One obtains $s \approx 0.822$ which gives $1-s \approx 0.175$. The risk sought after is therefore about 18%. Compare this with the risk of succeeding in 200 different tries,

$$\frac{200}{1024} \approx 0.195 = 19.5\%$$

9.3 Answers and Explanations to the Exercises in Section 3.3.7

1. Yes, the drone's family tree produces Fibonacci numbers generation after generation. This follows from the fact that the family tree pattern



corresponds exactly to the rabbit pair chronology (with time running vertically upward):

original pair — newborn pair

sexually mature pair

newborn pair

2. If we draw Figure 3.3.18b with a large scale and draw carefully, we find a narrow but obvious gap (space) between the top pieces (A + D) and the bottom pieces (B + C). The gap's shape is a parallelogram and the area is exactly r^2. The area A + B + C + D thus remains the same as in Figure 3.3.18a, $441r^2$. The gap is the "catch." How could this be discovered without drawing?

3. One first obtains

$$f_4 = 8 + 15 = 23,\ f_5 = 15 + 23 = 38, \ldots$$

$$f_9 = 259 \text{ and } f_{10} = 419.$$

Next,
$$\frac{f_9}{f_{10}} = \frac{259}{419} \approx 0.6181 \approx 0.618$$

4. We have
$$v_{n+1} = \frac{f_{n+1}}{f_{n+2}} = \frac{f_{n+1}}{f_{n+1} + f_n} = \frac{1}{1 + \frac{f_n}{f_{n+1}}} = \frac{1}{1 + v_n}$$

If the sequence $\{v_n\}$ is assumed to have a limiting value x, then the recursive formula
$$v_{n+1} = \frac{1}{1 + v_n}$$

leads to the fact that the limit x must satisfy the equation
$$x = \frac{1}{1 + x}$$

The limit must also be positive. The positive root of the equation above is precisely
$$G = \frac{1}{2}(\sqrt{5} - 1)$$

5. The proof is based on the following three points:

1) A monotonic decreasing sequence with a lower bound has a limit. This is an important theorem which we do not prove here.

2) $v_1 = \frac{1}{1} > G$ (whose value is 0.618...)

3. $v_{n+2} = \frac{1 + v_n}{2 + v_n}$

Relation 3) comes from the recursion formula we used above in Exercise 4:

$$v_{n+2} = \frac{1}{1 + v_{n+1}} \quad \text{in which we put} \quad v_{n+1} = \frac{1}{1 + v_n}$$

We now show that the numbers v_n with odd index form a monotonically decreasing sequence. Since they are bounded by zero, these numbers must have a limiting value, say A.

From 3) follows

$$v_n - v_{n+2} = \frac{v_n^2 + v_n - 1}{2 + v_n} = k_n\,(v_n^2 + v_n - 1) \qquad (4)$$

where k_n is positive.

If $v_n > G$ we have $v_n - v_{n+2} > k_n(G^2 + G - 1) = 0$ (see Section 3.3.5)

that is $v_n > v_{n+2}$.

In an analogous fashion we can show that $v_{n+2} > G$ (if $v_n > G$).

In other words, 2) implies that $v_1 > v_3 > v_5 > \ldots > G$

If $v_n < G$ we get instead, using the corresponding argument, that

$$v_n < v_{n+2} < G.$$

Since $v_2 = \frac{1}{2}$ we must have $v_2 < v_4 < v_6 < \ldots < G$

Applying 1) means that this sequence $\{v_{2n}\}$ too must have a limiting value, say B. Both limits A and B must according to relation 3) satisfy the following equation

$$x = \frac{1 + x}{2 + x}$$

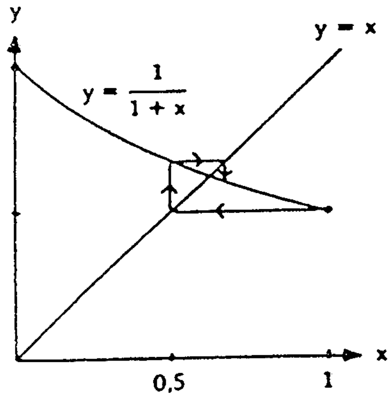
This equation gives $x^2 + x = 1$, i.e. has G as its positive root. We have thereby shown not only that the sequence $\{v_n\}$ converges to G but also that the numbers with odd index *decrease* toward G while numbers with even index *increase* toward G.

If we represent the recursive relation $v_{n+1} = \frac{1}{1 + v_n}$

graphically in x-y-coordinates with the aid of the curve $y = \frac{1}{1 + x}$

and the line $y = x$,

we get a picture of the convergence process (see the figure).



6. Starting with $F_1 = a$ and $F_2 = b$ gives

$$F_3 = a + b$$
$$F_4 = a + 2b$$
$$F_5 = 2a + 3b$$
$$F_6 = 3a + 5b$$
etc.

We see that the F-numbers take the form

$$F_n = f_{n-2}a + f_{n-1}b \text{ (for } n \geq 3)$$

where $f_1 = 1$, $f_2 = 1$, $f_3 = 2$, $f_4 = 3$, $f_5 = 5$ etc., are the "original" Fibonacci numbers we became acquainted with in Section 3.3.2.

a) We get $F_{10} = 21a + 34b$

b) If we let $v_n = \frac{F_n}{F_{n+1}}$, we get the same recursion formula as in Exercise 5, namely

$$v_{n+1} = \frac{1}{1 + v_n},$$

and the proof that the sequence $\{v_n\}$ converges to G works analogously to the proof in Exercise 5. From this it follows that the limit is independent of a and b.

7. We consider the second of the two given sequences,

$$G,\ 1,\ \frac{1}{G},\ \frac{1}{G^2},\ \ldots$$

From Section 3.3.5, equation (2), we know that G satisfies

$$G^2 + G = 1 \qquad (2)$$

If we divide this equation by G we get

$$G + 1 = \frac{1}{G} \qquad (3)$$

If we again divide by G we get

$$1 + \frac{1}{G} = \frac{1}{G^2} \qquad (4)$$

and thereafter in similar fashion $\frac{1}{G} + \frac{1}{G^2} = \frac{1}{G^3}$ and so on. (5)

The equations (3), (4), (5) etc., show that the sequence with G, 1, $1/G$, $1/G^2$, . . . is a Fibonacci sequence with starting values G and 1. It must therefore be identical to the sequence G, 1, G + 1, ...

8. If the regularity continues infinitely, then it can be expressed

$$f_{2n-1} = f_n^2 + f_{n-1}^2 \quad \text{for } n \geq 2 \qquad (1a)$$

$$f_{2n} = f_{n+1}^2 - f_{n-1}^2 \quad \text{for } n \geq 2 \qquad (1b)$$

The proof is by induction: we know that the above applies for f_3 and f_4 ($n = 2$). Assume that the formula applies for $n = p$.

We then get for $n = p + 1$

$f_{2p+1} = f_{2p} + f_{2p-1}$ by definition and according to (1a) and (1b)

$$f_{2p+1} = f_{p+1}^2 - f_{p-1}^2 + f_p^2 + f_{p-1}^2 = f_{p+1}^2 + f_p^2$$

In other words, the regularity also holds for f_{2p+1}.
It remains to be shown for f_{2p+2}.

$$\begin{aligned} f_{2p+2} = f_{2p+1} + f_{2p} &= \left(f^2_{p+1} + f^2_p\right) + \left(f^2_{p+1} - f^2_{p-1}\right) \\ &= 2f^2_{p+1} + f^2_p - f^2_{p-1} \\ &= 2f^2_{p+1} + f^2_p - (f_{p+1} - f_p)^2 \\ &= f^2_{p+1} + 2f_{p+1}f_p. \end{aligned}$$

From $f_{p+2} = f_{p+1} + f_p$ it follows (after squaring and rearranging items) that

$$2f_{p+1}f_p = f^2_{p+2} - f^2_{p+1} - f^2_p$$

Putting this expression into equality (2) now gives

$$f_{2p+2} = f^2_{p+2} - f^2_p$$

and thereby the applicability of (1b) is shown even for $n = p + 1$.

Since the equalities (1) and (1b) apply for $n = 2$ they must according to the above also apply for $n = 3$, for the same reason also for $n = 4$, etc. In other words, they must hold for all values of n.

9.4 Answers and Explanations to the Exercises in Section 3.4.7

1. The fourth term is 8.2, the fifth 10.6.

2. The fourth number is $12.8 \cdot \frac{8}{5} = 20.48$

The fifth term is $\frac{20.48 \cdot 8}{5} = 32.768$

3. After 3 follows $3 \cdot \frac{6}{12} = 1.5$

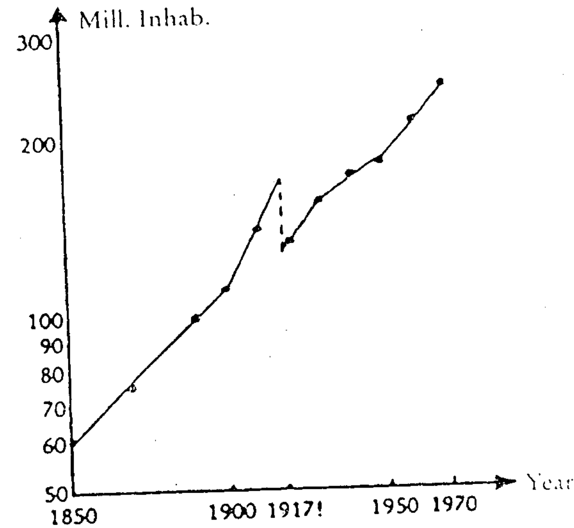
After that comes 0.75.

4. The angle of the sixth swing is $15^\circ \cdot \left(\frac{12}{15}\right)^5 = 15^\circ \cdot 0.8^5 \approx 4.9^\circ$.

6. The light *remaining* is $0.95^4 I$, where I is the intensity of the incoming light. The percentage of light *lost* is

$$100 \cdot (1 - 0.95^4) \approx 18.5\%.$$

7. The plot shows a discontinuity between 1910 and 1920. This comes from the Russian Revolution in 1917. The population growth is shown plotted below in a semi-log diagram.



9.5 Answers and Explanations to the Exercises in Section 3.5.8

1. $D - d = 2t$, $D = d + 2t$, $d = D - 2t$, $t = \frac{D - d}{2}$

2. Yes, the numbers which turn up there are identical with the numbers in Pascal's triangle (if we put a 1 at the top). Let us look at the case with 4 numbers to begin with: a, b, c, and d.

Starting point	1a	1b	1c	1d
We get first	1a + 1b	1b + 1c	1c + 1d	
Next comes	1a + 2b + 1c	1b + 2c + 1d		

from the addition

$$\begin{array}{ccc} 1 & 1 & \\ & 1 & 1 \\ \hline 1 & 2 & 1 \end{array}$$

And finally we get $1a + 3b + 3c + 1d$

from adding

$$\begin{array}{cccc} 1 & 2 & 1 & \\ & 1 & 2 & 1 \\ \hline 1 & 3 & 3 & 1 \end{array}$$

The numbers in Pascal's triangle turn up in a corresponding manner:

$$\begin{array}{ccccccc} & & & 1 & & & \\ & & 1 & & 1 & & \\ & 1 & & 2 & & 1 & \\ 1 & & 3 & & 3 & & 1 \end{array}$$

3. No, because if the square were

$$\begin{array}{cc} a & b \\ c & d \end{array}$$

then the following would hold:

$$a + b = a + c = a + d$$

from which it follows that $b = c$ and $c = d$.

In other words, the numbers would not be different. We would find instead that

$$a = b = c = d.$$

4. We found in Section 3.5.4 that three consecutive integers give a sum which is divisible by 3. If the "middle" number is n, then the sum will be

$$s_3 = (n - 1) + n + (n + 1) = 3n.$$

If we take four consecutive numbers, there is no "middle" number, but if we let the second number be n, then the sum will be

$$s_4 = (n - 1) + n + (n + 1) + (n + 2) = 4n + 2$$

These sums are *not* evenly divisible by 4 (there is a remainder of 2). Taking 5 numbers in a row we get

$$s_5 = (n - 2) + (n - 1) + n + (n + 1) + (n + 2) = 5n$$

which tells us these sums are all divisible by 5.

It is now not difficult to see the following general result:

1) when n is odd, the sum is divisible by n.
2) when n is even, say n = 2p, the sum gives a remainder of p, when divided by 2p

5. Let E represent an even number, T any number divisible by 3, and S any number divisible by both 2 and 3, i.e. divisible by 6. Finally, let O stand for an odd number which is not divisible by 3. Prime numbers greater than 3 can be found only among the O-numbers. Beginning with T, the sequence of natural numbers can be written in the following structured way:

(3)	(4)	(5)	(6) ...						(12)...			
T	E	O	S	O	E	T	E	O	S	O	E	T...

The period SOETEO repeats endlessly because

1) Every other number is even
2) Every third number is divisible by 3
3) Every sixth number is divisible by 6

The O-numbers therefore always lie adjacent to the S-numbers.

We have thereby shown that every prime number greater than 3 can be written in the form 6n + 1 or 6n - 1, where n is a natural number.

6. According to Exercise 5 the prime numbers can all be written in the form $6n + 1$ or $6n - 1$. Assume that $p = 6n + 1$, for some integer n. Then we get

$$p^2 + 2 = (6n + 1)^2 + 2 = 36n^2 + 12n + 3 = 3\,(12n^2 + 4n + 1)$$

from which it is seen that the number is divisible by 3.

If instead $p = 6n - 1$, we get analogously

$$p^2 + 2 = 3\,(12n^2 - 4n + 1)$$

and conclude as before that the number is divisible by 3.

7. We suppose that the formula

$$1^3 + 2^3 + 3^3 + \ldots + n^3 = (1 + 2 + 3 + \ldots + n)^2$$

holds for n = some natural number p.

We know, to begin with, that the formula holds for $n = 1$ (we have, of course, also checked it for a few other values of n). Does it hold now for $n = p + 1$? We need to utilize the knowledge that the right hand side in the formula above can also be written

$$\left\{\frac{n\,(n+1)}{2}\right\}^2 \quad \text{or} \quad \left\{\frac{n^2\,(n+1)^2}{4}\right\}. \qquad (^+)$$

We can now write, adding $(p+1)^3$ to both sides,

$$1^3 + 2^3 + \ldots + p^3 + (p+1)^3 = \frac{p^2(p+1)^2}{4} + (p+1)^3$$

We must now show that the right hand side can be written as

$$\{1 + 2 + \ldots + p + (p+1)\}^2.$$

The right hand side can be written

$$\frac{(p+1)^2}{4}\{p^2 + 4(p+1)\}$$
$$= \frac{(p+1)^2}{4}(p^2 + 4p + 4)$$
$$= \frac{(p+1)^2(p+2)^2}{4}$$

Comparison with formula $(^+)$ above shows that this expression is precisely

$$\{1 + 2 + \ldots + p + (p+1)\}^2.$$

We have thus shown that if the formula for the sum of cubes is true for $n = p$, then it is also true for $n = p + 1$, i.e., for the next higher value of n. Since the formula holds for $n = 1$, it must hold for $n = 2$, thus even for $n = 3$, and we have hereby proven the formula for all natural numbers n.

8. We let x be the amount of copper to be added. The equation for x will then be

$$0.6 + x = 0.7\,(1 + x)$$

which gives $\quad x = \dfrac{1}{3}$ kg.

9.6 Answers and Explanations to the Exercises in Section 3.6.4

1. The Danes own 2/7 of 50% = approximately 14.3% of Linjeflyg.

2. The increase was 24,000 crowns, which is $4 \cdot 6000$ or 400%.

3. Since 0.5% is 1/200, one must mine 200 tons of ore to get 1 ton of copper.

4. Letting p be the original price, the price after the 25% increase is 1.25 p crowns. The discount down to the original price will be 0.25 p crowns which as percentage becomes

$$\frac{0.25p}{1.25p} \cdot 100 = 20\%.$$

5. The 50% decrease means a halving of the contribution. The 100% increase means a doubling, so the school contribution ends up at the original level.

6. If we assume for simplicity's sake that the bones are circular in cross section, then the radius must be enough larger to multiply the area by a factor of 8. The radius must, therefore, be multiplied up by a factor $\sqrt{8}$ or approximately 2.8 times larger.

7. Average speed = total distance / total time.

$$\text{Total distance} = \frac{10}{60} \cdot 20 + \frac{30}{60} \cdot 15 = \frac{65}{6} \text{ nautical miles.}$$

$$\text{Total time} = 10 + 30 \text{ min} = 2/3 \text{ hour.}$$

$$\text{Average speed is thus } \frac{65/6}{2/3} = \frac{65 \cdot 3}{6 \cdot 2} \approx 16.25 \approx 16 \text{ knots.}$$

8. Let a be the distance travelled outbound in kilometers. We call the airplane in the problem airplane A. A's average speed is the same as the constant speed which another airplane B would need to hold in order to fly the same distance (out and back) in the same time as it took for A.

Let x be A's average velocity = B's constant velocity.

$$\text{A's travel time is } \frac{a}{800} + \frac{a}{1000} \text{ hours.}$$

$$\text{B's travel time is } \frac{2a}{x} \text{ hours.}$$

$$\frac{2a}{x} = \frac{a}{800} + \frac{a}{1000}.$$

The travel times are to be the same, which gives the equation

The fact that a can be divided out of the equation shows that the answer is not dependent on the distance.

$$\text{We get } \frac{2}{x} = \frac{5+4}{4000}$$

$$\text{from which we get } x = \frac{8000}{9} \approx 889 \text{ km/hour.}$$

9.7 Answers and Explanations to Section 3.7.3

1. Figure 9.7.1 shows that the cube has 11 different networks. These are ordered systematically: in the first row we have the 6 networks which have 4 squares in a row. In the second row are the four networks with 3 squares in a row. The eleventh network is the only one where no more than 2 squares lie in a row. The tetrahedron has only 2 possible networks.

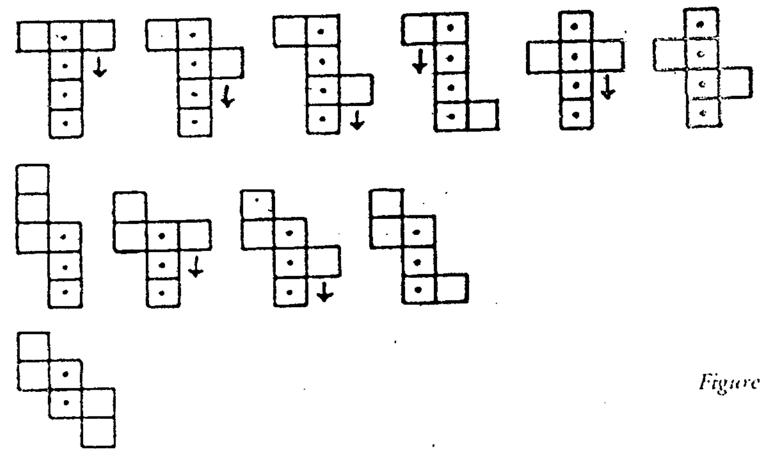


Figure 9.7.1

2. Finding eight corner points for the cube should not be difficult.

3. See Figure 9.7.2.

4. See Figure 9.7.3.

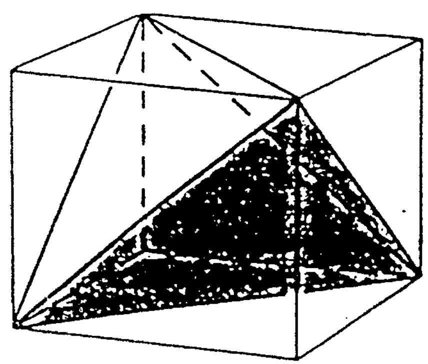


Figure 9.7.2

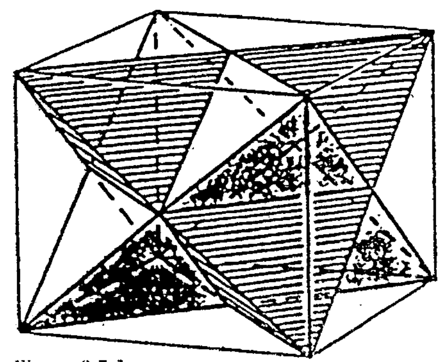


Figure 9.7.3

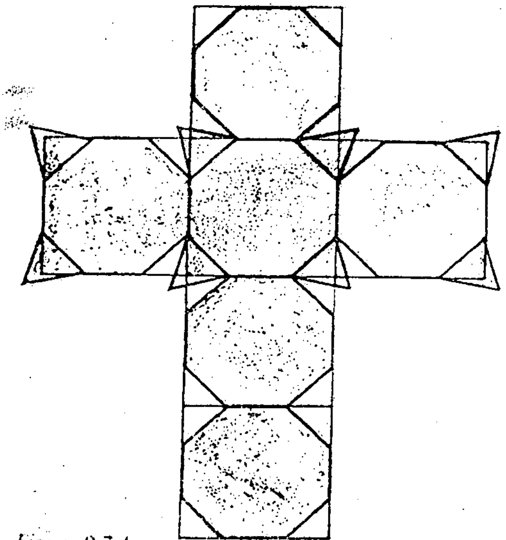


Figure 9.7.4

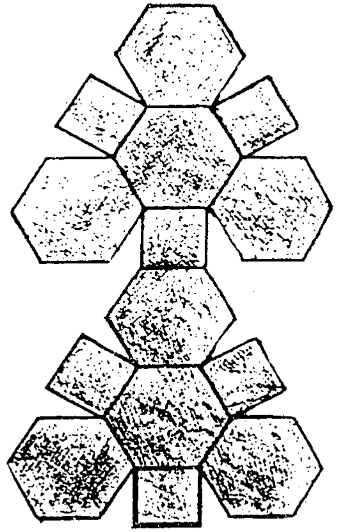


Figure 9.7.5

5. The network for the solid with octagons and triangles is easy to construct when one knows how to inscribe an octagon in a square: one constructs four quarter circles with centers at the corners of the square and with radius equal to half the diagonal of the square. The square has a side equal to the edge length chosen for the cube. See Figure 9.7.4.

The network for the solid with hexagons and squares is more difficult to find, since one must first figure out the length of the sides which form these figures. It is not too difficult to see in Figure 3.7.8 that the square's diagonal is equal to half of the side of the cube (see Figure 9.7.5). The network for this solid is noticeably smaller than that for the octagon-triangle solid. This is because we are grinding down the solid the whole time.

9.8 Answers and Explanations to Section 3.8.3

1. See Figure 9.8.1. The three curves correspond to XY = 5, 15, and 25 mm respectively.

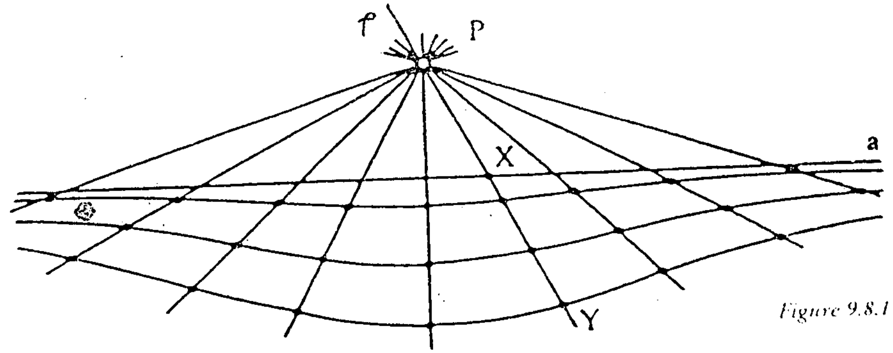


Figure 9.8.1

2. In order to get an idea of the shape of the curve for large XY values, we reduce the figure in size using so-called proportional form reduction with P as the fixed point: each p-line has a point X belonging to line a and a point Y belonging to the curve. The distance PX and PY are reduced to such a scale that the distance between the new points, X´Y´, is for example 25 mm. In this type of reduction (proportional form) all shapes and forms are unchanged, i.e., line a becomes a new line a´ which is closer to P and parallel to a. The curve too (the locus of points Y) will maintain its shape at the same time as it moves nearer to P.

For large values of the distance XY, a´ will lie very close to P. For example, if XY = 1 meter the reduction scale must be 1:40, which means that P lies 40 times closer to a´ than a.

We may now with the same effect imagine line a to be fixed and let P be drawn in toward a while we take XY to be 25 mm. Figure 9.8.2 shows the shape of the curve for the cases where P is at distances 5 mm and 1 mm from a, respectively.

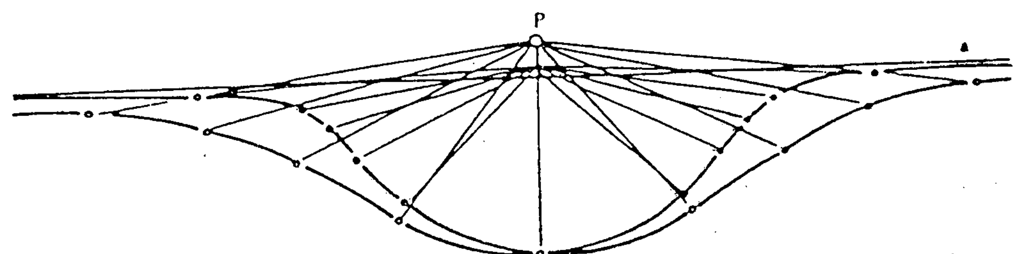


Figure 9.8.2

It is now not difficult to see that Figure 9.8.3 with its half-circle (or radius 25 mm) shows the limiting form of the family of curves as P is moved in toward a. To give a strict proof requires both care and "technique," the latter at university level.

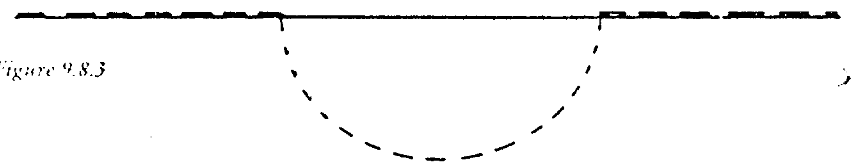


Figure 9.8.3

3. See Figure 9.8.4.

Figure 9.8.4

4. The 9 sub-figures in Figure 9.8.5 show the typical stages of the transformation. In (1) the three circles coincide but begin to move outward away from each other. Each circle's center follows its triangle "height" line (from the base at right angles up to the vertex) in the direction away from the vertex where the height line begins.

White region: not covered by any circle
Horizontally shaded regions: covered by 1 circle
Cross-hatched regions: covered by 2 circles
Dotted regions: covered by all 3 circles

In drawing (7) the circle radii have become infinitely large. In (8) the circles' inner regions lie "outside" the edge of the circle, in the sense that we used in Chapter 3.8. Drawing (9) shows a stage where we come back to (1) as far as form is concerned, but (9) and (1) are each other's opposites with regard to shading. It would take 9 more stages before we would truly come back to our starting point.

5. With reference to Figure 3.8.14 we let A and B be two given points and F a fixed proportional ratio between the distance from a point X on the curve to A and B respectively:

$$\frac{XB}{XA} = F$$

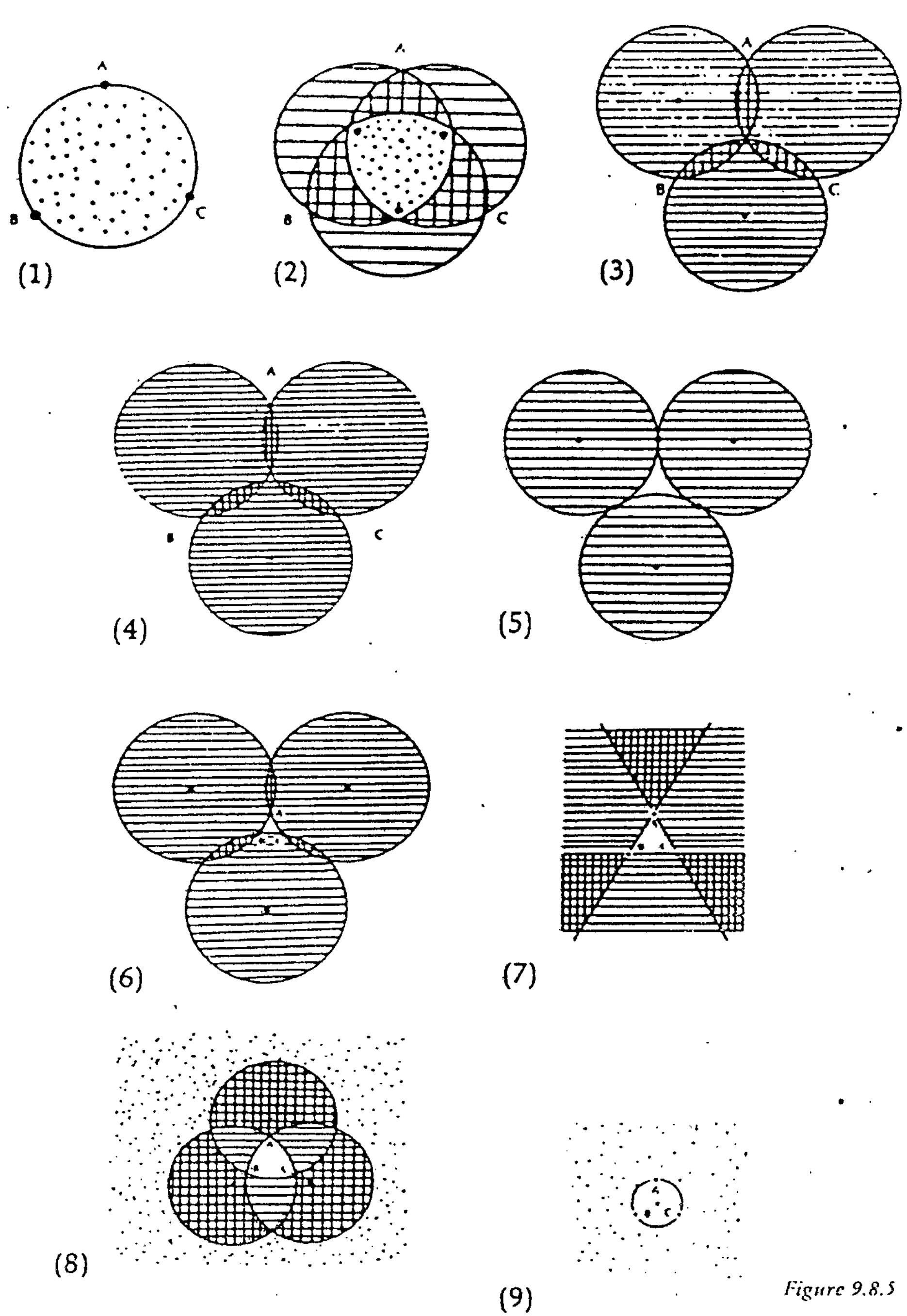


Figure 9.8.5

In Figure 9.8.6 below XP is drawn as a bisector in the triangle ABX. XQ is the bisector of the angle BXr, where r is the extension of AX. XQ is usually called the "external" bisector. According to the so-called Theorem of Bisectors and its correspondence for external bisectors, we have

$$\frac{PB}{PA} = F \quad \text{and} \quad \frac{QB}{QA} = F \qquad (1)$$

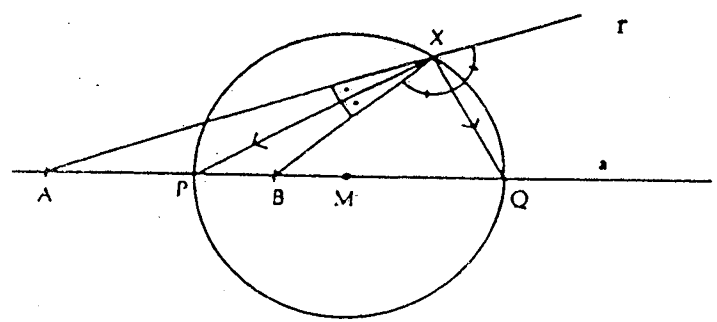


Figure 9.8.6

If we now let X move, then P and Q will be fixed points because of (1). In addition, the angle PXQ is, and always remains, a right angle (= half of 180°).

The conclusion must now be that X describes a circle with PQ as a diameter.

9.9 Answers and Explanations to Section 3.9.5

1. The distance to the North Pole is $\frac{(90-60) \cdot 40000}{360}$ km, or more directly $(90-60) \cdot 111.1$ km 3300 kilometers.

The distance to the South Pole is approximately 20000-3300 = 16,700 km.

2. The distance is s $(151-18) \cdot 111.1 \cdot \cos 34° \approx 12{,}250$ km.

3. The best answers are

a) approximately 9700 km (the angular difference is 87°)
b) approximately 12,500 km (angular difference 113°)

The solution to b) is illustrated in Figure 9.9.1 below.

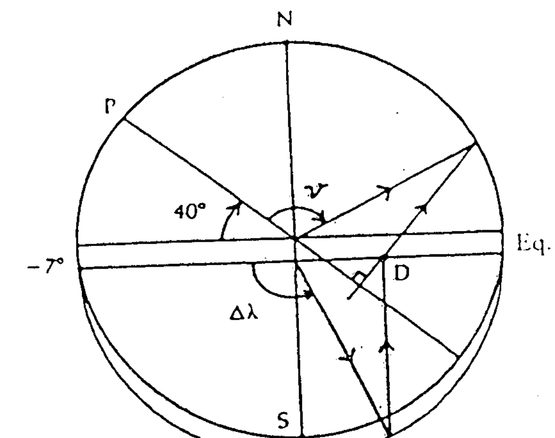


Figure 9.9.1

4. Because 106° + 76° = 180° it happens that New York and Hanoi lie on a great circle which goes through the Poles. The shortest distance passes through the North Pole (first toward the North, then southward) and the distance is

$$[(90 - 41) + (90 - 21)] \cdot 111.1 \approx 13000 \text{ km}.$$

5. Somewhere in the Southern Hemisphere there is a circle of latitude whose length is exactly 1000 km. The starting point can be anywhere 1000 km north of that circle. The person may even start 1000 km north of any circle of latitude whose length is

$$\frac{1000}{n} \text{ km},$$

where n = 1, 2, 3, 4, etc.

6. According to the formula for the area of a spherical triangle (see Section 3.9.2) we get the equation

$$\frac{\pi R^2}{180}(V - 180) = 0.25 \cdot 4\pi R^2$$

where R is the radius of the sphere. Solving, one gets V = 360°.

7. Letting X be the percentage we are looking for, we have

$$\frac{x}{100} \cdot 4\pi R^2 = 2.2\pi R\,(R \sin 60° - R \sin 30°)$$

from which x ≈ 36.6%.

9.10 Answers and Explanations to Section 3.10.3

The symbol $\overset{=}{\wedge}$ means "perspective with" and d denotes Désargues' line.

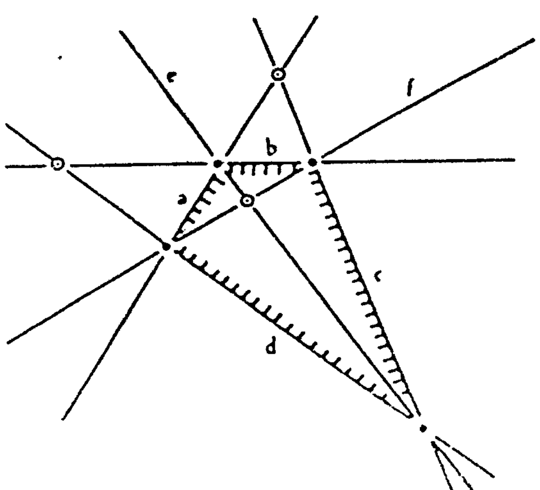


Figure 9.10.1

1. We obtain A´PR $\overset{=}{\wedge}$ OBC and d = B´C´Q

2. OAB $\overset{=}{\wedge}$ C´RQ; d goes through A´, B´and P.

3. a) AA´P $\overset{=}{\wedge}$ CC´Q; R is the center of perspective.

b) The figure is self-dual: on each line lie 3 points and through each point go 3 lines. The statements to be shown are consequences of this fundamental symmetry.

4. In Figure 9.10.1, the sides of the quadrangle are denoted by a, b, c, and d, while the two diagonals are labelled e and f.

The 3-sided polygon we are looking for is determined by the points of intersection

a x c, b x d and e x f.

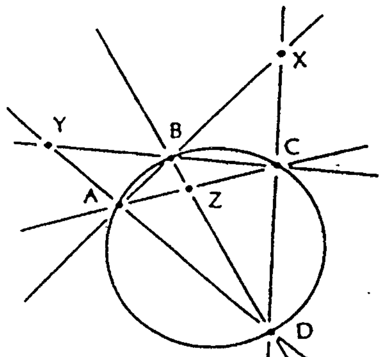


Figure 9.10.2a

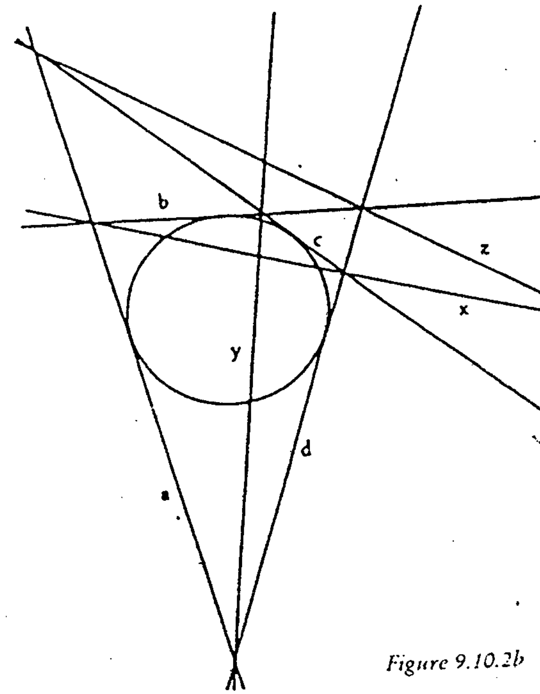


Figure 9.10.2b

5. Corresponding to three points X, Y, and Z in Figure 9.10.2a are the three lines x, y, and z in Figure 9.10.2b.

Z is an inner point of the circular area in Figure 9.10.2a. Analogously, z is an inner line in the circle envelope of lines which surround the circle in the b-figure. X and Y are outer elements, as are x and y.

6. See Figure 9.10.3

7. See Figure 9.10.4

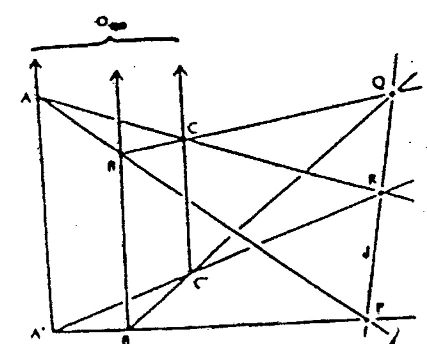


Figure 9.10.4
0_∞ is the center of perspective

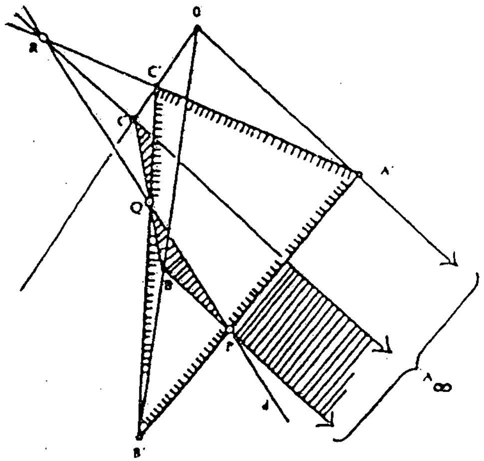


Figure 9.10.4
Triangle ABC has vertex A as a point at infinity

9.11 Answers and Explanation to Section 3.11.8

1. a) 21 pupils (if all those who were positive toward France were also positive toward Germany).

b) 5 pupils

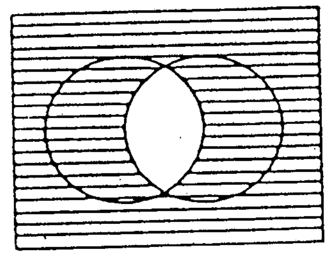
c) 16 pupils (= 25 + 21 - 30)

2. Assume for the sake of simplicity that the numbers of respondents was 100. Let a, b, and c represent the sets of respondents who have confidence in A, B and C respectively. (These sets have common elements, as is clear from the problem statement.)

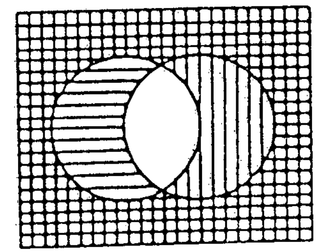
We are to minimize the intersection of a, b, and c. The minimum intersection of a and b is made up of 50 people (80 + 70 - 100). We now place as much as possible of c outside this intersection, which is 50 people. The remaining 10 c-respondents must be part of the intersection of a and b. In other words, the intersection of a, b, and c must include at least 10 people. The answer is 10%.

3. and 4.

The diagrams for $A' \cup B'$ and $(A \cap B)'$(and respectively for $A' \cap B'$ and $(A \cup B)'$) agree regardless of whether the sets overlap one another or not. The figures show $A' \cup B'$ and $A' \cap B'$ when A overlaps B. It can be seen that these sets are identical with $(A \cap B)'$ and $(A \cup B)'$ respectively.



$A' \cup B'$ = the entire shaded area. This corresponds to the complement of the remaining unshaded lens-shaped area, i.e., to $(A \cap B)'$.



$A' \cap B'$ = the cross-hatched area. This corresponds to the area outside the circles, i.e., to $(A \cup B)'$.

5. Simply follow the rules of dualization.

6. Let f be the left hand side = $x \cdot (y + z)$
and g be the right hand side = $x \cdot y + x \cdot z$

We must calculate f and g for all possible combinations of the values 0 and 1 for x, y and z and then show that f = g.

The table can be condensed down to

x	y	z	f	g
0	arbitrary		0	0
1	0	0	0	0
1	other comb.		1	1

Thus it is true that f = g.

The other identity can be verified with the table

x	y	z	f	g
0	1	1	1	1
0	other comb.		0	0
1	arbitrary		1	1

7. The table is

x	y	f	g
0	0	0	0
0	1	1	1
1	0	1	1
1	1	1	1

from which it is clear that f = g for all combinations of x and y.

9.12 Answers and Explanations to Section 3.12.8

1. The average speed is 129 km/hour.

2. The voltage increases from 2.0 mV to 4.8 V, an increase of 2.8 V.

The average rate of increase is $\frac{2.8\text{V}}{4.4\text{mA}} = 0.64$ V/mA.

3. The average speed is $\frac{1.6(4.5^2 - 2^2)}{4.5 - 2} = 10.4$ m/s

4. $\frac{(t+h)^2 - t^2}{h} = \frac{(2t+h) \cdot h}{h} = 2t + h$

and the limit is $2t$, when $h \to 0$.

5. If one doesn't know the Binomial Theorem for development of the coefficients of $(t + h)^n$,

then the easiest way is to set $u = t + h$

and consider the ratio $\frac{u^n - t^n}{u - t}$

The numerator can be rewritten

$$(u-t)\,(u^{n-1} + u^{n-2}t + u^{n-3}t^2 + \ldots + t^{n-1}).$$

The limit we seek is equal to the long expression in parentheses on the right, since the factor u - t can be divided out.

The limit $h \to 0$, that is, is $u \to t$, is

$$t^{n-1} + t^{n-1} + \ldots + t^{n-1} \quad (\text{n equal terms}) = n\,t^{n-1}.$$

6. Let us assume that $f(x) \neq 0$ at some point t within the interval. Since f (x) is continuous there must then be a subinterval where f (x) and thus $f(x)^2$ is non-zero. Let δ be the length of this closed subinterval and $m > 0$ be the minimum value of $f(x)^2$ on the subinterval.

Then, since $f(x)^2$ is non-negative, we have

$$\int_a^b f(x)^2 dx \geq m \cdot \delta > 0$$

which contradicts the given condition that $\int_a^b f(x)^2 dx = 0$.

7. a) One finds the primitive function

$$F(x) = \frac{x^2}{2} - \frac{x^3}{3}$$

and obtains $F(1) - F(0) = \frac{1}{6}$.

b) The area in question is

$$\int_2^3 x^2 dx = \frac{3^3}{3} - \frac{0^3}{3} = 9 \text{ cm}^2$$

(The primitive function is $x^3/3$.)

8. The statement's validity follows from the correspondence of pairs below:

$$1 = 1^2 \leftrightarrow 1$$
$$4 = 2^2 \leftrightarrow 2$$
$$9 = 3^2 \leftrightarrow 3$$
$$16 = 4^2 \leftrightarrow 4$$

etc.

9. The set of rational numbers a/b between 0 and 1 can be numbered in the following way (the list here shows the first 17 numbers):

$$\frac{1}{2};\ \frac{1}{3}, \frac{2}{3};\ \frac{1}{4}, \frac{3}{4};\ \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5};\ \frac{1}{6}, \frac{5}{6};\ \frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7};\ \ldots$$

The numbers are arranged in order of increasing denominator and within each group in order of increasing numerator. Fractions such as 2/4 and 2/6, 3/6, 4/6 are not included since they have the same value as a previously counted rational number.

10. Letting 0 represent the choice of a left branch (seen from below) and 1 be a choice to the right, then every path can be described by an infinite sequence of zeroes and ones. Each path has a unique representation.

We now copy Cantor's diagonal proof and assume that the infinite set of paths, represented as zero-one sequences, is countable. There would then exist an infinitely long list containing all the sequences, for example

1)	0001010...
2)	0110100...
3)	1010100...
4)	1001011...
....	

We now form a sequence where

the 1st digit ≠ 1st digit of 1);
the 2nd digit ≠ 2nd digit of 2)
etc.

The infinite sequence we obtain cannot possibly be on the list. We have contradicted our assumption that the list includes all sequences.

The paths are, therefore, uncountable.

BIBLIOGRAPHY

This bibliography does not claim to be as extensive as it could and perhaps should be. It contains a selection of works which I have met. Among the books there are surely numerous ones that are available in new editions with a later year of publication.

A. General References

1. Alexandrov, A. D., Kolmogorov, A. N., Lavrent'ev M. A.: *Mathematics, its Content, Methods and Meaning*, M.I.T. Press, Cambridge (Mass.) 1969.
2. Becker, O.: *Grundlagen der Mathematik in geschichtlicher Entwicklung*, Karl Alber, Freiburg/Br. 1954.
3. Bergamini, D.: *Mathematics.* LIFE - Wonders of Science, Time-Life International, Time Inc., New York 1969.
4. Courant, R., Robbins, H.: *What is Mathematics?* Oxford Univ. Press, Oxford 1981.
5. Davis, J. D., Hersh, R.: *The Mathematical Experience*, Birkhauser, Boston 1981, Penguin, Harmondsworth 1983.
6. Kline, M., *Mathematics in Western Culture*, Oxford Univ. Press, Oxford 1968.
7. Kline, M.: *Mathematical Thought from Ancient to Modern Times*, Oxford Univ. Press, Oxford 1972.
8. Locher-Ernst, L.: *Mathematik als Vorschule zur Geist Erkenntnis*, Philosophisch-Anthroposophischer Verlag, Dornach 1973.
9. Meschkowski, H.: *Didaktik der Mathematik*, Klett, Stuttgart 1972.
10. Newman, J. R.: *The World of Mathematics*, Simon & Schuster, New York 1956.
11. Otte, M. (Ed.): *Mathematiker uber die Mathematik*, Springer, Heidelberg 1974.
12. Polya, G.: *How to Solve It. A New Aspect of Mathematical Method*, Princeton Univ. Pr., Princeton, N.J. 1945.
13. Polya, G.: *Mathematical Discovery; on Understanding, Learning, and Teaching Problem-Solving*, 2 vols., Wiley, New York 1962-65.
14. Polya, G.: *Mathematics and Plausible Reasoning, 1. Induction and Analogy in Mathematics, 2. Patterns of Plausible Inference.* Princeton Univ. Press
15. Schuberth, E.: *Die Modernisierung des Mathematikunterrichts*, Verlag Freies Geistesleben, Stuttgart,1971.
16. Speiser, A.: *Klassische Stücke der Mathematik*, Orell Füssli, Zürich 1925.
17. Wagenschein, M.: *Ursprüngliches Verstehen und exaktes Denken*, Klett, Stuttgart 1963.
18. Wickelgren, W. A.: *How to Solve Problems*, Freeman, New York 1974.
19. Wittenberg, A. J.: *Bildung und Mathematik*, Klett, Stuttgart 1963.

B. References to Chapters 2 - 8

CHAPTER 2

Appenzeller, K.: *Die Quadratur des Zirkels*, Zbinden, Basel 1979.

Aaboe, A.: *Episodes from the Early History of Mathematics*, Yale Univ. Press, New Haven 1964.

Archimedes, *The Works of Archimedes with the Method of Archimedes*, (ed. T.L. Heath) Dover Publ., New York.

Dantzig, T.: *Number, The Language of Science*, Macmillan, New York/London 1954.

Heath, T.L. (Ed.): *The Works of Archimedes with the Method of Archimedes*, Dover Publ., New York.

Midonick, H.: *The Treasury of Mathematics*, Penguin, Harmondsworth 1965.

Newman, J. R.: "The Rhind Papyrus" (see A 10).

Polya, G.: See A 13 and 14.

Smith, D. E.: *History of Mathematics*, Dover Publ., New York 1958

Waerden, B. L. van der: *Science Awakening*, P. Noordhoff 1954.

Section 3.1

American Math. Monthly, vol. 83, 1976 (see 3.1.7).

Canadian Math. Bull. vol. 18/3, 1975 (see 3.1.7).

Coe, M. D.: *The Maya*, Penguin, Harmondsworth 1977.

Dantzig, T.: *Number, The Language of Science*, Macmillan, New York/London 1954.

Davis, M., Hersh, R.: "Hilbert's 10th Problem", *Scient. Amer.* 1973/11.

Davis, P. J.: *The Lore of Large Numbers*, Yale Univ. Press, New Haven 1961.

Freudenthal, H.: *Mathematik in Wissenschaft Und Alltag*, Kindler, Munchen 1968.

Gardner, M.: "The Remarkable Lore of Prime Numbers" *Scient. Amer.* 1964/3 and see also 1980/12.

Gardner, M.: "Simplicity as a Scientific Concept" *Scient. Amer.* 1963/8.

Locher-Ernst, L.: "Die Reihe der natürlichen Zahlen als Geist-Kunstwerk", *Sternkalender* 1959/60, Philosophisch-Anthroposophischer Verlag, Dornach.

Menninger, K.: *Zahlwort Und Ziffer*, 2. Aufl. Vandenhoeck & Ruprecht, Gottingen 1957.

Smith and Ginsburg: Article on numbers and digits in A 10.

Ulin, B.: "Beweis oder Gegenbeispiel?" *Mathematisch-Physikalischer Korrespondenz* Nr. 113, Dornach 1984.

Vogel, K.: *Vorgriechische Mathematik*, I und II, Schoningh, Paderborn 1958-1959.

McWhorter, E. W.: "The Small Electronic Calculator" *Scient. Amer.* 1976/3.

Section 3.2

Feller, W.: *An Introduction to Probability Theory and Its Applications*, Vol. 1,2, Wiley & Sons, New York 1965.

Gardner, M.: "The Multiple Charm of Pascal's Triangle:, *Scient. Amer.* 1966/12.

Gardner, M.: "Taxicab Geometry Offers a Free Ride to Non-Euclidean Locale", *Scient. Amer.* 1980/11.

Gnedenko, B. W.: *Lehrbuch der Wahrscheinlichkeitsrechnung*, Akademie-Verlag, Berlin 1965.

Huff, D.: *How to Take a Chance*, Norton & Company, New York 1964.

Locher-Ernst, L.: *Arithmetik und Algebra*, 2. Auflage, Philosophisch-Anthroposophischer Verlag, Dornach 1984.

Polya, G.: *Patterns of Plausible Inference*, (Vol. 2, A14). Princeton Univ. Press 1954.

Weaver, W.: *Lady Luck: The Theory of Probability*, Educational Services Inc. 1963.

Section 3.3

Baravalle, H. von: *Die Geometrie des Pentagramms und der Goldene Schnitt*, Mellinger, Stuttgart 1950.

Coxeter, H. S. M.: *Introduction to Geometry*, Chapter11, 2nd edition, Wiley & Sons, New York 1969.

Gardner, M.: *The 2nd Scientific Amer. Book of Math. Puzzles and Diversions*, Simon & Schuster, New York 1961.

Gardner, M.: "The Multiple Fascinations of the Fibonacci Numbers" *Scient. Amer.* 1969/3.

Locher-Ernst, L.: "Von der Geometrie der menschlichen Gestalt" (in A 8).

Morike-*Mergenthaler, Biologie des Menschen*, Quelle & Meyer, Heidelberg 1963.

Schultz, J.: "Die Blattstellungen im Pflanzenreich als Ausdruck Kosmischer GesetzmaBigkeiten" *Goethe in Unsrer Zeit*, Hybernia-Verlag, Dornack-Basel 1949.

Vorobyov, N.N.: *The Fibonacci Numbers*, Heath & Co., Lexington 1963.

Zeising, A.: *Neue Lehre von den Proportionen des Menschlichen Körpers*, Leipzig 1854.

Section 3.4

Bindel, E.: *Logarithmen fur Jedermann*, Verlag Freies Geistesleben, Stuttgart 1983.

Bindel, E.: *Die Zahlengrundlagen der Musik*, Verlag Freies Geistesleben, Stuttgart 1985.

Boring, E. G.: "Gustav Theodor Fechner" (Article in A 10, Chapter VI).

Locher-Ernst, L.: "Mathematik und Musik" (Article in A 8).

Malthus, T. R.: Article in A 10, Chapter VI.

Pfrogner, H.: *Lebendige Tonwelt*, Langen/Muller, Munchen 1981.

Stumpf, K. (Hrsg.): *Astronomie*, Das Fischer-Lexikon Bd. 4, Chapter "Helligkeiten", Fischer, Frankfurt/M. 1957.

Section 3.5

Becker, O.: *Grundlagen der Mathematik in Geschichtlicher Entwicklung*, Karl Alber, Freiburg/Br. 1954 (speziell Kapitel 2A).

Bindel, E.: *Die Arithmetik*, Mellinger, Stuttgart 1967.

Bindel, E.: *Das Rechnen*, Mellinger, Stuttgart 1966.

Gardner, M.: *aha! Insight*, Freeman & Co., New York 1978.

Schuh, F.: *The Master Book of Mathematical Recreations*, Dover Publ., New York 1968.

Section 3.6

Huff, D.: *How to Lie with Statistics*, Penguin, Harmondsworth 1973.

Physical Science Study Committee: *Physics*, Chapter "Functions and Scaling", Heath & Co., Lexington 1960.

Section 3.7

Baravalle, H. von: *Geometrie als Sprache der Formen*, Verlag Freies Geistesleben, Stuttgart 1980.

Bentley, W. A./Hunphreys, W. J.: *Snow Crystals*, Dover Publ., New York 1962.

Bindel, E.: *Harmonien im Reiche der Geometrie*, Verlag Freies Geistesleben, Stuttgart 1964.

Brun, V.: "On Some Problems in Solid Geometry Inspired by Virology" *Nordisk Matematisk Tidskrift* 1978/3 and 4.

Gardner, M.: "The Multiple Fascinations . . ." (see references for Section 3.3).

Haldane, J.B.S.: "On Being the Right Size" (Article in A 10).

Hilbert, D./Cohn-Vossen, S.: *Anschauliche Geometrie*, Springer, Berlin 1932, Dover Publ., New York 1944.

Schüpback, W.: *Pflanzengeometrie*, Troxler, Bern 1944.

Thompson, D'Arcy Wentworth: *On Growth and Form*, Cambridge Univ. Press, Cambridge 1963.

Unger, G.: *Das offenbare Geheimnis des Raume*, Verlag Freies Geistesleben, Stuttgart 1963.

Weyl, H.: *Symmetry*, Princeton Univ. Press, 1952.

Section 3.8

Adam/Buhler u. a.: *Beitrage Zum Geometrischen Zeichnen*, Verlag Freies Geistesleben, Stuttgart 1967.

Adams, G./Whicher, O.: *Die Pflanze in Raum und Gegenraum*, Verlag Freies Geistesleben, Stuttgart 1960.

Bindel, E.: *Die Kegelschnitte*, Verlag Freies Geistesleben, Stuttgart 1963.Bockemuhl, J. (Ed): Erscheinungsformen des Ätherischen, Verlag Freies Geistesleben, Stuttgart 1977.

Edwards, L.: *The Field of Form*, Floris Books, Edinburgh 1982.

Gordon/Jacobson: "The Shaping of Tissues in Embryos", *Scient. Amer.* 1978/6.

Grohmann, G.: *Metamorphosen im Pflanzenreich*, Verlag Freies Geistesleben, Stuttgart 1958.

Kranich, E. M.: *Die Formensprache der Pflanze*, Verlag Freies Geistesleben, Stuttgart 1976.

Locher-Ernst, L.: *Geometrisieren im Bereiche Wichtigster Kurvenformen*, Orell Fussli, Zurich 1938.

Strakosch, A.: *Geometrie durch übende Anschauung*, Mellinger, Stuttgart 1962.

Thompson, D'Arcy W.: see references for Section 3.7.

Wyss/Adam/Ruchti: *GTZ, Geometrisch-Technisches Zeichnen*, Lehrerhandbuch, Staatlicher Lehrmittelverlag, Bern 1978.

Section 3.9

Baldus, R./Löbell, F.: *Nichteuklidische Geometrie*, Sammlung Göschen, Bd. 970, Berlin 1953.

Balser, L.: *Einführung in Die Kartenlehre*, Teubner, Stuttgart 1951.

Locher-Ernst, L.: *Geometrische Metamorphosen*, Philosophisch-Anthroposophischer Verlag, Dornach 1970.

Zacharias, M.: *Das Parallelenproblem und Seine Lösung*, Teubner, Stuttgart 1951.

Section 3.10

Bernhard, A.: *Projektive Geometrie Aus der Raumanschauung zeichnend Entwickelt*, Verlag Freies Geistesleben, Stuttgart 1984.

Cremona, L.: *Elements of Projective Geometry*, Dover Publ., New York 1960.

Gardner, M.: "Euclid's Parallel Postulate and Its Modern Offspring", *Scient. Amer.* 1981/10.

Kline, M.: *Mathematics in Western Culture*, Kapitel 10, Oxford Univ. Press, Oxford 1968.

Locher-Ernst, L.: *Urphänomene der Geometrie und Projektive Geometrie*, Orell Füssli, Zurich 1937 and 1940.

Locher-Ernst, L.: *Raum und Gegenraum*, Philosophisch-Anthroposophischer Verlag, Dornach 1957, 2. Aufl. 1970.

Rovida, Andes A., *Übungen Zur Synthetischen Projektiven Geometrie*, Verlag am Goetheanum (Schweiz) 1988.

Bell, E. T.: *Men of Mathematics*, Simon & Schuster, New York 1937.

Boole, G.: *An Investigation of the Laws of Thought*, Dover Publ., New York.

Descartes, R.: *Philosophical Works of Descartes*, Dover Publ., New York 1955.

Gardner, M.: *Logic Machines, Diagrams and Boolean Algebra*, Dover Publ., New York 1968.

Hilbert/Ackermann: *Grundzüge der theoretischen Logik*, Dover Publ., New York 1946, 6. Aufl. Springer, Heidelberg 1972.

Midonick, H.: *The Treasury of Mathematics, II*, Penguin, Harmondsworth 1965.

Newman, R. J.: see reference A 10, Chapters XI-XIV.

Whitesitt, J. E.: *Boolean Algebra and Its Applications*, Addison-Wesley, Reading (Mass.) 1961.

Wright, G. H. von: *Logik, filosofi och språk*, Aldus/Bonnier, Stockholm 1957.

Section 3.12

Apostol, T. M.: *Mathematical Analysis*, Addison-Wesley, Reading (Mass.) 1974.

Bell, E. T.: *Men of Mathematics*, see references in Section 3.11.

Goursat, E.: *Cours D'analyse Mathématique*, Paris 1943.

Klein, F.: *Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert*. Vol. 1 and 2., Springer, Heidelberg 1979.

Meschkowski, H.: *Probleme des Unendlichen Werk und Leben Georg Cantors*, Vieweg, Braunschweig 1967, 2 Edition as Georg Cantor. *Probleme des Unendlichen*. Bibl.=Inst., Mannheim 1983.

Meschkowski, H.: *Wandlungen des mathematischen denkens*, Vieweg, Braunschweig 1956, edition #4, 1967.

Niven, I.: Numbers: *Rational and Irrational*, Yale Univs., Pres, New Haven 1961.

Reid, C.: *A Long Way from Euclid*, Crowell & Co., New York 1963.

Waismann, F.: *Einführung in das mathematische denken*, Deutscher Taschenbuch-Verlag, München 1970.

CHAPTER 4

Chintschin, A. J.: "Über den Formalismus im Mathematikunterricht der Schule", *Probleme des Mathematikunterrichts*, Volk und Wissen, Berlin 1965.

Gardner, M.: *Aha! Insight*, Freeman & Co., New York 1978.

Kline, M.: *Why Johnny Can't Add: The Failure of the New Math*, St. Martin's Press, New York 1973.

Locher-Ernst, L.: *Mathematik als Vorschule Zur Geist-Erkenntnis*, Philosophisch-Anthroposophischer Verlag, Dornach 1973.

Newman, J. R.: Chapters XI-XIV in A 10.

Poincaré, H.: Article on mathematical creativity. in A 10.

Polya, G.: See A 12-14.

Waerden, B. L. van der: *Einfall und Überlegung*, Birkhäuser, Basel 1968.

Werrheimer, M.: *Produktives Denken*, Frankfurt/M. 1964.

CHAPTER 5

Andrade, E. N. da C.: *Sir Isaac Newton*, London 1954.

Burtt, E. A.: *The Metaphysical Foundations of Modern Physical Science. A Historical and Critical Essay*, Routledge & Kegan Paul, London 1924, 1972.

Caspar, M.: *Kepler*, Kohlhammer, Stuttgart 1958.

Descartes, R.: See reference in Section 3.11.

Dijksterhuis, E. J.: *Die Mechanisierung des Weltbildes*, Springer, Heidelberg 1956.

Dunnington, G. W.: *Carl Friedrich Gauss, Titan of Science*, Exposition Press, New York 1955.

Finsler, P.: "Formale Beweise un die Entscheidbarkeit" *Math. Zeitschr.* 25 (1926) 676-682.

Finsler, P.: *Aufsätze Zur Mengenlehre*, Wiss. Buchgesellschaft, Darmstadt 1975.

Godel, K.: "Über Formal Unentscheidbare Sätze der Principia Mathematica und verwandter Systeme 1" *Monatshefte F. Math und Phys.* 38 (1931) 173-198.

Hemleben, J.: *Galilei*, Rowohlt, Reinbek 1969.

Newman, J. R.: Chapter I in A 10.

Newton, I.: *Mathematical Principles of Nature*, Berlin, 1872.

Pascal, B.: *Aufsätze*, ausgewählt und eingeleitet von R. Schneider, Fischer 1954.

Stegmüller, W.: *Unvollständigkeit und Unentscheidbarkeit.* Die Metamathematischen Resultate von Godel, Church, Kleene, Rosser und ihre erkenntnistheoretische Bedeutung, Springer, Heidelberg 1973.

Steiner, R.: *Geisteswissenschaftliche Impulse Zur Entwicklung der Physik.* Erster naturwissenschaftlicher Kurs. Zehn Vorträge gehalten in, Stuttgart vom 23. 12. 1919 bis 3.1 1920. Gesamtausgabe Bibl.-Nr. 320, Rudolf Steiner Verlag, Dornach 1964.

Unger, G.: "Carl Friedrich Gauss" *Sternkalender* 1977/78, Philosophisch-Anthroposophischer Verlag, Dornach.

Valery, P.: *Les pages immortelles de Descartes*, Ed. Correa, Paris 1941.

Williams, L. P.: *Michael Faraday*, Chapman & Hall, London 1965.

CHAPTER 6

Albers/Alexanderson (edit): *Mathematical People*, Birkhäuser, Basel 1985.

Dijksterhuis, E. J.: See references for Chapter 5.

Glöckler, G.: "Zur Frage des Gebrauchs von Elektronischen Taschenrechnern in der Schule" *Erziehungskunst* 1977/6 u. 7, Verlag Freies Geistesleben, Stuttgart.

Heath, T. L. (Ed.): *The Works of Archimedes with the Method of Archimedes*, Dover Publ., New York.

Heitler, W.: *Der Mensch und die naturwissenschaftliche Erkenntnis*, Vieweg, Braunschweig 1961. 1962.

Hilbert, D.: *Grundlagen der Geometrie*, 12. Aufl., Teubner, Stuttgart 1977.

Reid, C.: *Hilbert*, Springer, Heidelberg 1970.

Piaget, J.: "How Children Form Mathematical Concepts" Scient Amer. 1953/11.

Wagenschein, M.: See reference.

Valéry, P.: *Les pages immortelles de Descartes*, Ed. Correa, Paris 1941.

CHAPTER 7

Meschkowski, H.: *Didaktik der Mathematik*, II, III, Klett, Stuttgart 1972.

Piaget, J.: *Die Entwicklung des Erkennens, I: Das mathematische Denken*, Klett, Stuttgart 1975.

Piaget, J.: "How Children Form Mathematical Concepts" *Scient. Amer.* 1953/11.

Steiner, R.: *Die geistig-seelischen Grundkräfte der Erziehungskunst* (Vortrag vom 25. 8. 1922), Gesamtausgabe Bibl. -Nr. 305, Rudolf Steiner Verlag, Dornach 1956.

Wagenschein, M. See reference A 17.

CHAPTER 8

Carleson, L.: "Matematik för vår tid" 1968, Stockholm (Swedish).

NAME INDEX.

Name, Section, *Page*

SUBJECT INDEX

Subject, Section, *Page*

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S