# WRITING MATHEMATICS MATH 370 

## Contents

0. The Course Topics ..... 2
Part 1. Writing Mechanics ..... 2
0.1. The standard steps in writing ..... 2
0.2. Example: writing about derivatives ..... 3
Part 2. Critical Thinking $\supseteq$ Science $\supseteq$ Mathematics ..... 4
1. Critical Thinking ..... 4
1.1. Paradoxes ..... 4
1.2. Lack of definitions in language and mathematics. ..... 7
2. The scientific method ..... 8
2.1. Effectiveness and limitation of the scientific method. ..... 8
3. The mathematical method ..... 10
3.1. Observation ..... 10
(A) Observations of the "outer" world. ..... 10
(B) Observations in the "inner" world. ..... 11
3.2. Abstraction ..... 12
3.3. Why definitions? Why proofs? ..... 13
4. Branches of mathematics ..... 13
Part 3. A two step example of the abstraction approach: Symmetries and groups ..... 15
5. Set Theory ..... 16
5.1. Set theory for groups ..... 17
6. Symmetries ..... 17
7. Groups

## $\bigcirc$

## 0. The Course Topics

- The writing technique
- The methods of: Critical Thinking $\supseteq$ Science $\supseteq$ Mathematics
- The latex typesetting program
- Group project with a joint classroom presentation
- Job search writing (Cover letter, resume, i.e., CV)
- Essays related to mathematics and mathematical careers


## Part 1. Writing Mechanics

0.1. The standard steps in writing. One can proceed systematically:
(1) Goals. One should formulate clearly what is the goal of the write up to be commenced. This includes a choice of audience (readership) for which the writing should be appropriate.
(2) The list of main elements of the story.
(3) Outline. It spells the progression of pieces that constitute the write up. So, it describes the structure of what you are about to write.
(4) Writing. This step is very personal. You will do it in your own way.

If one has a block one can try "free writing" to some extent. I mean that you write anything that comes to mind and is clearly related to the goals of the text. Once you write down a few such ideas you try to get more organized. See what is missing and say whatever you can about that. Then save this version as file " 1 " and copy it to a new file called " 2 ".

Now try to organize the bunch of sentences you have in file "2". This means to move pieces around, expand them and also prune the things for which you now have a better approach.
(5) Editing. Once you have produced some early version of the text you want to edit it. The criteria are something like this:

- Can a person in my target audience follow my story?

An exercise in handling this question is peer review.

- Is it interesting to me?

Another question of this type: is it likely to be interesting to my target audience?

- Is something wrong with the form of this writing?

For instance, do I spend too much time on something that contributes little to the text?

- Eliminate technical errors such as miss-spelling (computer will help with this).
(6) Rethinking and Rewriting. If you come back to the text after a little break (hours or days) you are likely to immediately spot some imperfection. (So, really here is always room for improvement.)
0.2. Example: writing about derivatives. We start with the goals. Let us say that you want to make incoming freshman appreciate the idea of a derivative. However, you do not want to copy the calculus textbook, rather you want to tell the story of derivatives from the point of view of someone who has a higher perspective derived from working through a number of courses in mathematics and maybe physics etc.
Some elements that could appear in your story:
(1) What do derivatives do for us.
(2) What is the intuitive idea of a derivative. $\left(f^{\prime}(a)\right.$ is the rate of change of $f(x)$ at $x=a$.)
(3) What is the historical origin of the idea in physics? (Velocity.)
(4) What is the geometric meaning of a derivative? (The slope of the tangent line.)
(5) What is the precise definition of the derivative $f^{\prime}(a)$ ? (The limit of the average rate of change of $f(x)$ in a small interval around $a$ as the size of the interval gets smaller and smaller.)
(6) Why is the formal definition reasonable?
(7) How does one calculate derivatives of functions in practice? (Using rules that reduce differentiation of a complicated function $f(x)$ to differentiation of simpler functions that $F(x)$ is a combination of. Using table of derivatives of elementary functions.)
(8) Some applications. (Minima and maxima, optimal design, ...)
(9) Example(s) of things like: (i) calculation of derivative by its formal definition; (ii) calculation of derivative by rules and tables; (iii) an application of derivatives.
(10) How much can be done with computers or scientific calculators?

One principle of effective writing: the question (1) ("what will you gain from reading this?"), should generally be tackled reasonably early since otherwise the reader may give up (not sure whether this will be useful) or find it difficult to follow (without knowing early which practical things the write up should be connected to). ${ }^{(1)}$

[^0]
## Part 2. Critical Thinking $\supseteq$ Science $\supseteq$ Mathematics

## 1. Critical Thinking

By critical thinking I mean that when you read, hear or see something, you at the same time try to process it, i.e., establish its relation to your experience. This involves evaluating the value of this information and what is its essential content.
The basic example for us will be the statements called paradoxes. These are stories which create confusion as some fact seems both true and false. The first use for us is that paradoxes give an opportunity to exercise critical thinking. The goal is to resolve the confusion by finding its source - some hidden inadequacy, failing or fallacy in the story.
We will in particular emphasize how those paradoxes demonstrate Imprecision and Richness of human language. The first element will be found to be the most obvious cause of the confusion in created by a paradox. The second element will manifest as deep and important messages hidden in the vehicle of a paradox story presented through the imperfect medium of our language.
1.1. Paradoxes. A. Liar's paradox. A man says "I am now lying". Try to decide whether what he says is false (F) or true (T) ?
One can easily see that the discussion of two possibilities T and F easily eliminates both. If "True" then he is now lying so what he is saying is false, hence "False". So, "True" is impossible since it implies (leads to) its opposite. Similarly, If "False" then he is now not lying, so he is telling the truth and we conclude that correct option for his statement is "True". So, "False" is also impossible since it implies its opposite.

The simplest way to resolve the paradox is to say that his statement - being neither True nor False - does not have a truth value. ${ }^{(2)}$

So, the observation is that language does not always work as precisely as we may imagine.
B. Barber paradox. In a village Santo Domingo in Spain there lived a barber who shaved all man in the village who did not shave themselves. Did he shave himself?

Again, we can eliminate both Yes and No. We see that this is just another version of liar's paradox. What does this mean? This time the problem is less abstract and the answer is simpler - there can exist a situation like that.
C. Russell's paradox. Let $S$ be the set whose elements are all sets $A$ that do not contain themselves, i.e.,

$$
S=\{A \text { is a set; } A \notin A\} .
$$

Does $S$ contain itself?

[^1]Again, both $S \in S$ and $S \notin S$ are both impossible. However, this time we do not find a contradiction in language but a contradiction in mathematics. Actually, Russell has intentionally translated liar's paradox into mathematics.
Where did we go wrong? What happened is that we believed in the following principles:
(1) Any set $X$ can be described by describing what are its elements, i.e., by giving a rule $\mathcal{R}$ that decides whether an object $x$ is an element of $X$ or not.
(2) Any such rule $\mathcal{R}$ defines a set $X$, its elements are all objects $x$ that satisfy the rule $\mathcal{R}$.

For instance in Russell's paradox, the "Russell rule"

$$
A \text { is a set and } A \text { is not an element of itself" }
$$

has been used to define the set $S$.
The mechanism that got us in trouble is that principle (1) is valid while principle (2) is not. Principle (1) is just the definition of sets: "a set is a class (collection) of objects that can be described by specifying which objects are its elements". Principle (2) is something that we just assumed should be valid because it seems to go well with the definition of sets. However, Russel's paradox shows that (2) is wrong, If (2) were correct then we could define the set $S$ as above by Russell's rule, but this leads to a contradiction.

Historical context. Set Theory is not very deep - mostly it is just a language and a notation that give us a way to describe mathematical objects. However, at the beginning of $20^{\text {th }}$ century mathematics encountered problems with some ideas that were clearly important but resisted precise definitions. The way they were used would produce many deep insights but occasionally also theorem that just turned out to be wrong. The way out of this crisis was Set Theory because it lead to very precise definitions of mathematical objects. (From this point of view any mathematical object should be describable as a combination of sets.)
Set Theory was at the time heralded as "paradise into which Cantor ${ }^{(3)}$ has brought mathematics". In 1901 Bertrand Russell cast doubt on this view on mathematics by showing the Set Theory is wrong - it contains contradictions.

Resolution of the paradox. The morale of the paradox is that if sets are meaningful objects, not any rule can be used to define a set - some rules are apparently meaningless. To save the Set Theory, a precise criterion on which rules define (meaningful) sets was introduced, called ZermeloFraenkel axioms of Set Theory. A much nicer explanation came much later. The point was that sets are not enough to describe mathematical objects, one needs to replace sets by a more general notion of categories which are collections of objects together with some information how these objects relate to each other. ${ }^{(4)}$

[^2]So, the slogan from the beginning of the $20^{\text {th }}$ century "All mathematical objects are (combinations of) sets." has worked quite well for a while but eventually had to be improved.
D. Zeno's (Xenon's) paradox. A hare and a turtle race 100 meters. The rabbit moves 10 meters per second and the turtle one meter per second. He is clearly much faster so he gives the turtle an advantage - he start when the turtle is already 10 meters into the race. At the moment when the hare makes up these 10 meters the turtle has moved an additional meter. When the hare makes up this 1 meter, the turtle has moved by 0.1 meters. And so on. So, we see that the ra-bit can never catch up with turtle - by the time he passes the remaining distance the turtle has moved even further so it always stays in front of the rabbit.

This story of course contradicts our intuition. Elementary physics says that the time when the rabbit will catch up is $\frac{10}{9}$ seconds and the place where they meet will be $\frac{100}{9}$ meters from the start.
The confusion in this story comes from its use of two notions of "time". Regarding the usual time (the number of seconds), the rabbit will catch up in $\frac{10}{9}$ seconds. The illusion comes from breaking the process of catching up into infinitely many steps and referring to the number of steps accomplished so far as "time". In the first step of this process rabbit advances by 10 meters (in 1 second), in the second it advances 1 meter in 0.1 seconds, then it advances 0.1 m in 0.01 seconds etc. The error happens when the story refers to the number of steps accomplished so far as time. This is hidden in the use of the word never, when we really mean "not after any finite number of these steps".
To see that the two ways of telling the story are compatible we just notice that the "catching up time" from the infinitely many steps process point of view is $1+0.1+0.01+$ $0,001+\cdots=1.1111 \cdots$ and this is the same as $\frac{10}{9}$.
Though it was easy to find the fallacy in this story, the story itself is not nonsense. It discovers the strategy of thinking of finite quantities in terms of infinite processes. Its paradoxical nature comes from the wonder of meeting such new idea.
This idea, new at the time, is actually the basis of all calculus. As we have seen in the differential calculus the derivative $f^{\prime}(a)$ can only be understood as a limit result of infinitely many calculations of average rate of change over smaller and smaller intervals.
In integral calculus the basic idea $\int_{a}^{b} f(x) d x$ means the area under the graph of the function $y=f(x)$ between $x=a$ and $x=b$. A priori we only know how to calculate directly the area of a rectangle. For a curved region $\mathcal{R}$ (for instance the region beneath $y=f(x)$ ), we can only directly calculate the are of a sub-region $\mathcal{R}^{\prime} \subseteq \mathcal{R}$ consisting of several rectangles. Then the actual area of $\mathcal{R}$ can only be understood as the limit of infinitely

[^3]many calculations of subregions $\mathcal{R}_{n}$ of $\mathcal{R}$ which consist of more and more rectangles in order to approximate better the region $\mathcal{R}$.
E. General Relativity and Quantum Mechanics. "Quantum Mechanics" is a part of physics that explains all forces of nature except gravitation. By "General Relativity" one means Einstein's explanation of how gravity functions. These are two highest achievements of physics. Each is an incredibly precise theory - it predicts the results of experiments with accuracy of more then 20 significant digits. ${ }^{(5)}$

However, there is a paradox in physics - these two theories contradict each other. So they can not be both true. Actually, neither is expected to be really true.
This is not as bad as it sounds. The goal of any science is to understand and describe some aspect of the world. However, this goal is never completely achieved - no science makes serious claims on really understanding the world. At any given time, the goal is more modest and still quite useful. One strives to produce an approximate description, which is sufficiently correct, i.e., up to a degree of precision that suffices for applications that we are interested in at that time.

The resolution of the "paradox" here is that Quantum Mechanics is very precise on the small scale of elementary particles. General Relativity is equally successful on the large scale of stars and planets. The fact that they contradict each other (when applied to the same scale) just means that we need a new idea which will be accurate on all scales and will behave like QM on the small scale and as GR on the large scale. The only candidate for such idea at the moment is called String Theory. ${ }^{(6)}$
F. Conclusion. Paradoxes present an opportunity for Critical Thinking. More importantly, they point to imprecisions in language or even in mathematics (bad), but they also often point to puzzles that will have a great impact when properly resolved.
1.2. Lack of definitions in language and mathematics. Here we notice that in the standard use of language lack of clear definitions of the words in general use is a cause of much confusion and conflict. The flip side is that one can only have the clear definitions for objects of systems that one understands. So, the absence of precise definitions in a given subject matter may also signal a depth and a potential of the subject.
Below, thinking about religion is used as an example of critical thinking. One reason that this is an interesting example is that it is a part of the most ambitious human attempt

[^4]to understand the world. Another reason is that it is sometimes a contentious topic so we have a particular opportunity to attempt to put aside the related passions and try to apply critical thinking. ${ }^{(7)}$
A. The meaning of "or" Here we notice that there are two (logically very different) uses of the same word "or". "While you are visiting you can play music or video games." means either the first or the second or both. In "Live free or die!" one means first or the second but not both (exclusionary "or"). In contrast, notation and terminology in mathematics and computer science make great difference between the two.
B. Define God. God should be something above our understanding. So it is quite reasonable that mankind has not yet come up with a satisfactory definition. What is used in practice is some partial list of attributes such as "God has created this world". ${ }^{8)}$

On the negative side the lack of definition here causes confusion (many people rely on their own understanding of what God is, to the extent of believing themselves to be above the others). On the other hand, lack of precise definitions is a natural consequence of ambitious goals.
C. Heat of political terminology. The observation here is that much of the basic lexicon of politics is often miss used. Technically, this means that the common use of communist, socialist, islamist, christian, liberal, conservative; does not come with a generally accepted definitions. Each of this phrases has a long history that gives the phrase a number of different meanings and emotional connotations. On a low level of discourse this is used to attack somebody without having to make clear why, or to make oneself look good without having to make your positions clear.

## 2. The scientific method

2.1. Effectiveness and limitation of the scientific method. I will try to express some successes and limitations of the scientific method by the following statement:

In the total of human experience science deals with relatively simple questions. With questions that it can be applied to, it is very effective.
The science is always wrong. The goal of science is not to produce a true description/explanation of the world, but to give approximate explanations that will be sufficiently correct for certain applications.

[^5]Here, "relatively simple" means that the topics that the scientific method applies should ideally be well defined and the claims that are made can be checked by some method, for instance an experiment.

Let us use terminology "hard sciences" for sciences that can use both methods and "soft sciences" for sciences for which this is less true.

Example. Hard sciences include mathematics, physics, chemistry etc. For instance consider the part of medicine that tries to come up with a drug that will cure certain illness. The candidate substance is administered to a group of sick animals or people. The effect is measured. If a sufficiently large percent gets better the substance is pronounced a new drug. If not, one tries to adjust it based on the results of the experiment.
This is a well defined sequence of steps. The shape of the process is determined though success may require huge amount of work, ingenuity or originality.

Example. If you study history you generally can not do experiments, you can not design a new war in order to check whether some theory of what works well in a war is correct. If you study or practice psychology you may have difficulty to come up with precise definitions of basic notions such as happiness, neurosis, schizophrenia, depression. (For example, the definitions of the American Psychiatric Society change in time.)

Example. In theology of philosophy, it is harder and often seemingly impossible to produce precise definitions of relevant objects.
In comparison, the progress in "harder" sciences seems faster because one can take the full advantage of the scientific method. On the other hand, "softer" sciences may be more ambitious and this is likely to make progress slower. (For instance it seems more complicated to understand why humans behave as they do then what will happen when two balls collide at given speeds and at a given angle.)
E. Science and religion. Here is what I come up with when I try to apply Critical Thinking to the topic of the relation of science and religion:

## Science does not apply to questions of religion.

The impression expressed here is that the basic ideas and claims of religion are not defined precisely enough so that the scientific method can be applied. For instance since we can not define god we can not discuss existence of god with any scientific accuracy.
Notice that a lack of clear definitions in the religious approach to the world need not be a sign of weakness but of strength when it is a measure of how ambitious the religious undertaking is: it attempts to deal with questions that are vastly too complicated for sciences.

So, roughly, the two approaches to the world actually do not share enough common ground to have great effect on each other. ${ }^{(9)}$

In fact, when scientific method and critical thinking apply to clashes between adherents of science and religion, it only seems to happen in a fairly negative way. These mechanisms can be used to disprove extreme claims of the "lost ones" or " bad apples" on both sides. When a religious leader makes claims on understanding specific aspect of the world based on his connection to god, say: "moon is painted on the sky to give us light at night"; such claim can be proved to be wrong by science (telescopes, moon trips). When a scientist claims that he understands the world better than he does: "I know how man has been created by evolution and therefore there is no place for god and so I know that there is no god"; then critical thinking disproves this line of thinking (Understanding mechanisms of evolution does not contradict role of god. God may have created evolution in order to produce a man. ${ }^{(10)}$ )

## 3. The mathematical method

I will emphasize the following aspects of the mathematical method of understanding things.
(1) Observation of the world as a source of mathematics.
(2) Abstraction as the key method of mathematics.
(3) Definitions and proofs as the language of mathematics.
3.1. Observation. Mathematics arises from observations of the world.

Here, I will (somewhat artificially) divide these in two classes

- (A) Observation of the "outer" world.
- (B) Observation of the "inner" world.


## (A) Observations of the "outer" world.

3.1.1. Motion and calculus. For instance observations of motion were abstracted into ideas of physics. The problem of understanding specific motion as it happens in time was then answered through mathematical type of thinking by developing calculus differentiation and integration.

[^6]3.1.2. Gravity and Differential Geometry. Albert Einstein found a very abstract explanation of what causes the gravitational attraction, called General Relativity. This required certain amount of abstract mathematics and his discover also spurned the development of Differential Geometry, one of the principal tools of contemporary mathematics.
3.1.3. Elementary particles and Quantum Field Theory. A more modern example is the observation of elementary particles. The physics that attempted to explain the behavior of elementary particles is the Quantum Field Theory. Much of mathematics in the last century was developed for purposes of solving problems in Quantum Field Theory.
(B) Observations in the "inner" world. Much of the present day mathematics has come to exist without a direct inspiration from the "real world" outside us. Rather, the mathematicians would work on abstract questions that they found interesting for reasons hard to explain, sometimes described as pursuit of ideas for their "inner beauty".
Just as everybody else, the practitioners of this abstract mathematics were convinced that their musings are of no practical importance. So, it appeared as a miracle when the ideas developed for their own sake were (often much later) found to be essential for understanding various sciences and also creating applications we use every day.
I will outline this mechanism of "the inner world" happening to be related to the "outer world" in the example of a branch of mathematics called Number Theory.
3.1.4. Number Theory. The basic objects in number theory are the integers. No mystery here, integers do come from the real world problems such as counting apples. However, after human mind found integers it was not satisfied with just using integers. It also started to ask various questions about integers themselves.

The simplest such questions were related to applications. For instance, finding an integer solution of $a^{2}+b^{2}=c^{2}$, such as $a=3, b=4, c=5$ was a way to draw the right angle in the ancient Babylon. However, the Babylonians then asked

What are all integer solutions $a, b, c$ of this equation?
This was already an abstract question of no interest in everyday life. The Babylonians actually found all solutions (the ancient Greeks were the first to write the solution clearly and explain why there are no more solutions). Since there are infinitely many genuinely different solutions, finding them all was a spectacular intellectual feat of humanity.
3.1.5. Fermat's conjecture. Fifteen (or more) centuries later (in 1637), a French judge Pierre Fermat, read a book on Greek mathematics and jotted in the margin:

For any exponent $n>2$ the equation $a^{n}+b^{n}=c^{n}$ has no solution in positive integers

$$
a, b, c .
$$

He also wrote that "I have a really beautiful proof but there is not enough space on this margin to write it down".

30 years later someone found these jottings and many people tried to reinvent Fermat's proof of his claim. It took next 100 years to supply the proof for $n=3$ : "there are no solutions of $a^{3}+b^{3}=c^{3}$ with $a, b, c$ positive integers". It took two more centuries until the general case (for any $n>2$ ) was proved (by Sir Andrew Wiles in 1994).
3.1.6. The effect of the study of Fermat's conjecture. Much of the mathematics known today has been discovered for the purpose of studying Fermat's claim. This includes some of the most important parts for applications to the "real world" such as the theory of groups which is essential in contemporary physics ("symmetries").
3.1.7. Number Theory and coding. The ideas of Number Theory have appeared in everyday use with the development of computers and internet. The security of electronic financial transactions is based on advanced tools of Number Theory, a science that has been developed for thousands of years with no real world applications. ${ }^{(11)}$

### 3.2. Abstraction.

3.2.1. The method of abstraction. It is concerned with ideas that appear in more than one setting. For instance, in the first semester of the algebra course (411) one studies the idea of a group and in the second semester (412) the idea of a ring. ${ }^{(12)}$ We will consider the process of abstraction in mathematics on the example of symmetries and groups in part 3.
For an idea like this, the process that we go through is

- Step 0. Spot an idea/mechanism that appears in several settings.
- Step 1. Name it.
- Step 2. Formulate this idea clearly.
- Step 3. Study it.
- Step 4. Apply it.

The gain is that whatever you have learned about this abstract idea in the third step now applies to all settings where this idea occurs.

[^7]3.2.2. Step 3. The study of an idea. To start with observe that our thinking is largely visual. This serves us well in subjects with visual intuition like geometry, but abstract thinking requires some getting used to. The approach is to combine two mechanisms:

- Examples. Whenever we get lost in abstract thinking we look for "down to Earth" examples of what this means.
- Developing the abstract thinking. We ask questions of interest about the idea we study. Since the object of study (certain idea) is abstract, answering such question affirmatively requires giving a clear explanation of why something is true, that's called a proof.

With a little exercise in both methods we get familiar with the abstract mathematical idea that we are considering.

Remark. This process is really how baby learns a language - a new word is acquired based on how it is used in examples and then words are combined into sentences.
So, developing abstract thinking (for instance abstract mathematics) takes us back to the learning games of early childhood.

In particular, in my experience we all have this ability. I have seen this through teaching math to a number of people, Also, there is a sheer fact that your brain has already excelled at this activity - once upon time we have all learned to speak. ${ }^{(13)}$

### 3.3. Why definitions? Why proofs?

3.3.1. Definitions. Mathematics differs from other "hard" sciences in that it deals with ideas. One can not directly see number $\sqrt{53}$, rather we form an idea of a number and introduce some notation (here: " $\sqrt{53}$ ") to represent this idea. A definition just means describing precisely what is the abstract idea that we are talking about.
3.3.2. Proofs. One of the features of handling abstract ideas is the notion of a "proof". This just means that in order to convince ourselves of an abstract fact requires an abstract explanation.

Remark. The careful use of definitions and proofs is a particular feature of mathematics. As we have seen, the reason mathematics is forced to use these tools is because it deals with abstract ideas so it can not rely on tools that other sciences use such as physical inspection of objects and experiments.

## 4. Branches of mathematics

[^8]4.0.3. Formulas. Much algebra is done by manipulating formulas. This is a very successful technical method. However, this is often not the crucial part, rather one needs to understand the ideas behind the formulas we write.

## Part 3. A two step example of the abstraction approach: Symmetries and groups

4.0.4. The first floor of our house: symmetries. Here, we follow the above steps of the abstraction method in order to distill from the examples in the "real world" the idea of a symmetry of a system. We make symmetry a mathematical object by finding its precise definition. Then we study this precise notion of symmetry using both examples and abstract considerations of properties of symmetries. We also consider the most standard application of symmetries: if a problem has symmetries then it can be simplified.
4.0.5. The second floor: groups. Now that the mathematical notion of a symmetry has becomes a familiar part of the "inner world" we use it as a setting for the the second use of the abstraction method. The observations of symmetries show that two symmetries of a given system $X$ can always be combined ("composed") to yield another symmetry. By considering how this composition functions we arrive at the notion of a group. Then we study groups and their applications in order to complete the second cycle of use of the abstraction method.

In this way on the edifice of the notion of symmetries (the ground floor of a house) we raise the second floor of the house - the theory of groups.

This process continues in time - since mathematics is cumulative humanity keeps adding the new floors.
4.0.6. Symmetries. In the introductory section 6 we study the notion of symmetry which is crucial in mathematics, physics and parts of chemistry. We start with an intuitive notion of a symmetry of an an object $X$, this notion is extrinsic in the sense that involves observers who look at $X$. Then we replace this extrinsic notion with an intrinsic notion of symmetries of $X$ - this one only involves the object $X$ itself, From this point of view a symmetry of $X$ is a motion of pieces of $X$ that preserves all relevant features of $X$.

For us the crucial object is the set $S_{X}$ of all (intrinsic) symmetries of an object (i.e., system) $X$. It turns out that the set $S_{X}$ of symmetries of an object has a useful structure: any two symmetries $f, g$ can be composed into a new symmetry $g \circ f$. Moreover, this structure has some important properties: associativity, neutral element and reversibility ("inverses").
4.0.7. Groups. The formal definition of the mathematical notion of a group appears in section 7 . This notion is an abstraction of structures and properties of symmetry sets $S_{X}$. A group is a set $S$ with an operation $*$ that combines two elements $a, b$ of $S$ into a single element denoted $a * b$. This operation is required to satisfy the above properties of the composition operation in sets of symmetries $S_{X}$.
4.0.8. Actions of groups on sets. The next step in this line of thinking (the third floor) would be the notion of actions of groups on sets. To arrive at this idea we we have to recall that the symmetries $S_{X}$ of any system $X$ "act" on the object $X$ - they moving around the pieces) of $X$. The abstract version of this idea is the notion of actions of groups on sets.
This third floor would complete the classical philosophical aspects of the notions of symmetries and groups.
4.0.9. Exposition of the theory of groups. In order to study groups it is not strictly necessary to start with the idea of symmetries. One can start directly with the more abstract style - the axiomatic definition of a group that here appears as the crown achievement. This "formal" definition is how groups are often introduced in books. However, starting with a formal definition may obscure what the groups are about, why they are natural objects to consider. and how they relate to more familiar "everyday" objects.

## 5. Set Theory

5.0.10. The word theory. It is often used differently in science than in everyday life. In its ordinary use it means a group of ideas that are formed in order to give a possible explanation of something that we do not understand well. In science "theory" means a large body of knowledge and understanding that has been developed over time and usually by a large number of scientists. In the first case the meaning is "maybe things work like that" and in the second "we have established that things work like that" and hence "we can present our ample evidence to anybody willing to check the claims of this theory". ${ }^{(14)}$
So, phrases like "set theory" or "group theory" mean a body of knowledge about idea of sets or or groups.
5.0.11. Set theory. A set is a mathematical word for a collection of things. The theory of sets is very simple, it systematically talks of what we can do with collections of objects.
Its basic advantage organizational and notational: the set theoretic notations allow us to replace some parts of sentences with a few simple symbols.
A deeper usefulness comes from the fact that all mathematical objects can be described in terms of sets - as systems of related sets. ${ }^{(15)}$ So, what the language of sets does for us is that it allows us to describe highly sophisticated (and therefore somewhat abstract, mysterious and confusing) mathematical objects precisely.

[^9]There is also a disadvantage. Presenting mathematical objects in terms of sets may seem unreasonably long winded and off putting to many.
5.1. Set theory for groups. The remainder of this subsection will only be (mildly) useful for our study of groups.

- Sets, Products of sets, Functions.
- Composition of functions. One first applies to $S$ the "closer" motion $g$ to get a point $g(S)$, and then one applies the "farther" motion $f$ to the result $g(S)$, and this produces $f(g(S))$. (Here "closer" and "farther" refer to the symbol $(f \circ g)(S)$ - in this symbol $g$ is closer to $S$ and $f$ is farther away.)
- Injections, surjections, bijections.

Lemma. A function has an inverse iff it is a bijection.

## 6. Symmetries

The mathematical notion of a group comes from the familiar idea of symmetry.
6.0.1. Objects as systems. We will consider symmetries of objects. The objects that we are interested in are systems consisting of smaller pieces. So, for such object $X$ the fist relevant quantity is the set $P_{X}$ of pieces of $X$. The next level of understanding $X$ is how these pieces should fit together to form object $X$. This will involve some "gluing information" telling us how the pieces fit.

Example. If $X$ is some geometric shape, say a triangle $T$ we can think of $T$ as consisting of all points that lie on $T$. From this point of view the pieces of $T$ are simply the points of $T$. The extra data is then the information on the distance $d(x, y)$ of all points $x, y$ in $T$. It tells us how to put together the points of $T$ to recover the triangle $T$ itself.

More generally, $X$ can be any "system" that consists of smaller pieces. It could be a machine, or water in a glass (consisting of molecules of water) or a single atom consisting of elementary particles (electrons, protons ...).
6.0.2. The intuitive (exterior) idea of symmetry. I will say that an object $X$ has a symmetry if it looks the same to two observers at different positions.

Example. Let $X$ be a square in the plane. Let $C$ be the center of the square. Consider two observers positioned at the point $P$ and $Q$ in the plane. If $Q$ is obtained by rotating $P$ by $90^{\circ}$ about the center $C$ of the square $X$ (in the counterclockwise direction), then the square will look the same to the two observers.

Example. A circle has infinitely many symmetries - one can rotate the observer about the center for any angle and from the new position it will look the same.
6.0.3. Symmetry of $X$ as a motion of pieces of $X$. Let us place a read light at a point $S$ in the square. When observer at $P$ takes a snapshot of the square the picture will in particular show the position of the point $S$. Suppose he emails the picture to the observer at $Q$ (and he also switches off the light at $S$ ). The observer at $Q$ now finds a point $S^{\prime}$ on the square which from his point of view corresponds to the red light in the picture.
So, we have found a procedure to sens a point $S$ on the square to another point $S^{\prime}$ which is in the same position for the observer at $Q$ as $S$ is for the observer at $P$. This means that we have created a motion of the square $X$ into itself that takes any point $S$ in $X$ to some point $S^{\prime}$.
Actually, there is no mystery - this motion is just the rotation of the square $X$ itself by $90^{\circ}$ about the center $C$ of the square $X$ and in the counterclockwise direction. ${ }^{(16)}$

Conclusion. Any symmetry of an object $X$ corresponds to some motion $f$ of pieces of $X$ which takes the whole object $X$ into itself.

Remark. A "motion of pieces of $X$ " really means a function $f: P_{X} \rightarrow P_{X}$ from the set $P_{X}$ of pieces of $X$ into itself - the value of the function $f(S)$ on a piece $S$ tell us that the piece $S$ has been moved into the place formerly occupied by another piece $f(S)$. So, symmetries are really functions.

### 6.0.4. Composition of symmetry motions.

Example. Rotations of a square. Let us denote by $\rho_{\alpha}$ the rotation of the square by the angle $\alpha$ (counterclockwise). It moves each point $S$ in the square to some point that we denote $\rho_{\alpha}(S)$.
Let us rotate the square by $90^{\circ}$, and then do one more rotation by $180^{\circ}$. So, we first apply $\rho_{90}$ and then to the result we apply rotation $\rho_{180}$. A point $S$ in the square has been moved to some point $S^{\prime}=\rho_{90}(S)$ by the first motion $\rho_{90}$. Then $S^{\prime}$ was moved to another point $S^{\prime \prime}=\rho_{180}\left(S^{\prime}\right)$ by the second motion $\rho_{180}$. The total effect of two motions is that $S$ was moved to $S^{\prime \prime}$.

This new motion is called the composition of two motions, it is denoted $\rho_{180} \circ \rho_{90}$. The way it acts on the point $S$ is by the rule

$$
\left(\rho_{180} \circ \rho_{90}\right)(S)=S^{\prime \prime}=\rho_{180}\left(S^{\prime}\right)=\rho_{180}\left(\rho_{90}(S)\right)
$$

In general for motions $f, g$ of pieces of $X$ the formula for the composed symmetry is simple - the effect $(f \circ g)(S)$ of the composition $f \circ g$ on a point $S$ is given by the rule (as in the above example):

$$
(f \circ g)(S) \stackrel{\text { def }}{=} f(g(S))
$$

[^10]So, this is just the composition of functions $f: P_{X} \rightarrow P_{X}$ and $g: P_{X} \rightarrow P_{X}$.
Conclusion. For any two motions $f, g$ of an object $X$ into itself there is another motion called the composition $f \circ g$ of two motions $f$ and $g$. This motion is the combined effect of first applying $g$ and then applying $f$.
If we think of motions $f, g$ as functions on the set $P_{X}$ of pieces of $X$ then the composition of symmetries is just the composition of functions.
6.0.5. Reversibility. This is an important property of motions that come from symmetries: such motions can be reversed.

In our example, rotating observer at $P$ by $90^{\circ}$ counterclockwise to position $Q$ has produced a motion $f$ of the square which works in the same way (counterclockwise $90^{\circ}$ rotation). If we now rotate $Q$ by $90^{\circ}$ in the opposite direction (clockwise) back into the position $P$ the corresponding motion of the square will be rotation $\rho_{-90}$ by $90^{\circ}$ clockwise. It reverses $\rho_{90}$ in the sense that when $\rho_{90}$ takes $S$ to $S^{\prime}$ then $\rho_{-90}$ takes $S^{\prime \prime}$ back to $S$.
Notice that this can be stated in terms of the composition: if we apply $\rho_{-90}$ after $\rho_{90}$ the total motion is the composition $\rho_{-90} \circ \rho_{90}$. the claim that $\rho_{-90}$ reverses $\rho_{90}$ means that the composition sends each point $S$ to itself, i.e.,

$$
\left(\rho_{-90} \circ \rho_{90}\right)(S)=S
$$

Conclusion. (i) Any object $X$ has one obvious motion - the motion of $X$ into itself which fixes all pieces of $X$. We will introduce the notation $i d_{X}$ (the identity of $X$ ) for this motion, so

$$
i d_{X}(S)=S
$$

for any piece $S$ of $X$.
(ii) The motions $f$ that come from symmetries of $X$ can be reversed, i.e., there is another motion $g$ such that

$$
g \circ f=i d_{X}=f \circ g
$$

Remarks. (0) The meaning of these two equalities is that what one of the motions does the other one reverses (we also say "cancels" or "undoes"), i.e., if say $f$ sends $S$ to $S^{\prime}$ then $g$ sends $S^{\prime}$ back to $S$ :

$$
g\left(S^{\prime}\right)=g(f(S)=(g \circ f)(S)=S
$$

(1) Motion $g$ which reverses $f$ is called the inverse of $f$ and it is denoted $f^{-1}$.
6.0.6. The set $\mathcal{S}_{X}$ of symmetries of $X$. Recall that objects $X$ that we consider a "systems" that consists of smaller pieces. Moreover, these pieces should fit with each other in a certain way. For instance any machine $X$ consists of many pieces that are put together in a certain way. The same for computer program.

Now we can give a meaningful definition of what we mean by symmetry motion of $X$. This is a motion $f$ of pieces of $X$ which can be reversed and which preserves the relevant relations between pieces of $X$.

We denote by $S_{X}$ the set of all symmetries of $X$.
6.0.7. What we mean by a system (or an object). The following examples should illustrate how important it is to understand precisely what mean by a given system $X$.

Example. (1) Let $L$ be the number line considered as a rigid object, i.e., we remember all distances between points of $L$. This is an example of a system $X$ - its pieces are the points of $L$ and the relevant information for two points on $L$ is their distance. So, $S_{L}$ consists of all motions $f$ of $L$ into itself which preserve the distance between points - for any points $a, b$ of $L$ the distance $d(a, b)$ between them is the same as the distance $d(f(a), f(b)$ between their $f$-images $f(a)$ and $f(b)$.
For instance, a translation by number $\alpha$ is a symmetry $T_{\alpha}$ of $L$, it acts by $T_{\alpha}(x)=x+\alpha$ for any $x \in L$. Also, for any point $a$ in $L$, the reflection $R_{a}$ in the point $a$ is a symmetry of $L$. If $R_{a}$ takes $x$ to $x^{\prime}$ then $a$ is the midpoint between $x$ and $x^{\prime}$, so $a=\frac{x+x^{\prime}}{2}$. Therefore, $2 a=x+x^{\prime}$ and $x^{\prime}=2 a-x$. So, $R_{a}(x)=2 a-x$.

Lemma. $S_{L}$ consists precisely of translations and reflections.
Proof. It remains to show that any motion $f$ in $S_{L}$ is either a translation or a reflection. We will postpone the (easy) argument for this until we get some abstract understanding of the nature of symmetry sets $S_{X}$ (see ??).

Example. (2) Now let $\mathcal{L}$ be the number line considered just as a set. (So, the distances are no longer a part of the system that we consider.) Now $f(x)=x^{3}$ is a symmetry of $\mathcal{L}$ - it movers around the points of $\mathcal{L}$ and it can be reversed - the inverse symmetry is $g(x)=\sqrt[3]{x}$. However, this $f$ is not a symmetry of the above system $L$ since it does not preserve the distances.

Example. (3) Let $A$ be a set and $\mathcal{P}_{A}$ be its partitive set, i.e., the set of all subsets of $A$. We can think of a system $X$ which consists of the set $P_{A}$ and the inclusion relation. So, for subsets $X, Y$ of $A$ we will remember the information of whether $X \subseteq Y$.

Lemma. The symmetries $S_{X}$ of the system $X$ are the same as symmetries of the set $A$.
Proof. Postponed to ??.
6.0.8. The algebraic structure of sets $S_{X}$ of symmetries of objects $X$. We have noticed that for any $X$ the set of its symmetries $S_{X}$ has the following structure

- (0) a composition operation $\circ$ which combines two symmetries $f$ and $g$ into a symmetry $f \circ g$.
- (2) a canonical element $i d_{X}$ which is "neutral" with respect to composition - for any $f$ in $S_{X}$

$$
f \circ i d_{X}=f=i d_{X} \circ X
$$

In other words, when we put $i d_{X}$ in a composition with $f$ it does not affect the symmetry $f$. ${ }^{(17)}$

- (3) Any symmetry $f \in S_{X}$ can be reversed, i.e., there is some $g \in S_{X}$ such that $f \circ g=i d_{X}=g \circ f$.

There will be one more property (1) called associativity, which we have not noticed yet.
It will turn out that these properties of $S_{X}$ are very useful. So, in 7 we will define the notion of a group as as any set with structures and properties listed in (0)-(3).
6.0.9. Symmetries as motions: the moral of the story. The new idea of viewing symmetries of $X$ as motions of $X$ into itself being

## 7. Groups

[^11]
[^0]:    ${ }^{1}$ This is not so essential if you have a captive audience.

[^1]:    ${ }^{2}$ Much more thought can be out into This is actually not unusual. If he said "Red dog moon yes" we would easily agree that this combination of words does not have a truth value. What caused confusion is an imprecision in language: a skillful combination of words may sound meaningful even if it is not.

[^2]:    ${ }^{3}$ Mathematician George Cantor has created the original ("naive") Set Theory.
    ${ }^{4}$ One actually also needs 2 -categories: one considers collections of objects but one also remembers how these objects relate to each other and how relations between objects relate to each other. Also,

[^3]:    3 -categories and 4-categories etc. In the end the correct all encompassing notion is that of $\infty$-categories where one remembers objects (called 0-morphisms), relations between 0-morphisms (called 1-morphisms), relations between 1-morphisms (called 2-morphisms), and so on forever. Clearly, this is a language which allows study of more complicated systems that contain complicated interdependence. Currently, the pure mathematics is undergoing the impact of using these more powerful ideas.

[^4]:    ${ }^{5}$ Significant digits start at the first nonzero digit, say .00032 has two significant digits.
    ${ }^{6}$ String Theory is sometimes called "a theory of everything" since it should give an explanation of "all known forces". However, what is meant is really "all forces that have been observed in nature until very recently". In the last decade new phenomena were found, dark matter and dark energy, for which we have no explanation at the moment.

    The main idea of String Theory is that the elementary particles are not tiny balls but tiny strings (closed into circles or just a piece of a string with two ends). Then the difference between elementary particles is explained as the difference in the way the strings are vibrating!

[^5]:    ${ }^{7}$ It should be clear that I am here not interested in what is your personal relation to religion. Also, none of this is meant to hurt anybody's feelings or change anyone's view at all. The logical analysis bellow finds no faults with religion or science, it just tries to discern some basic strengths and limitations of both methods.
    ${ }^{8}$ One may claim that our understanding of the world is paltry. An argument for this is that in the last 2 or 3 thousand years it has been changing drastically (now even within a single person's lifetime), and there seems to be no reason to expect that this will not continue. So, how could we imagine that we understand its creator?

[^6]:    ${ }^{9}$ Here, "roughly" means that there is some indirect interaction. For instance a claim has been made that a belief in a single god (monotheism) spurned the science in the West because it suggested that world makes sense and therefore one could try to understand it. Also, modern physics paints a much more rich picture of the world than the Newtonian physics. Again, claims have been made that this new picture of the world is more alike the one offered by religion.
    ${ }^{10}$ It is even theoretically possible that the world was created 3000 years ago but was made to look as if it is billions of years old. At the beginning of a movie the characters look as if they have existed for many years.

[^7]:    ${ }^{11}$ There seems to be a trend - the more the abilities of the human race grow, the larger the role of mathematics.
    ${ }^{12}$ Actually, this mathematical terminology does not reflect the usual meaning of these words in English! So, there is a definition of the meaning of the expressions group, ring in mathematics.

[^8]:    ${ }^{13}$ The belief that some of us have, that they "can not do math", seems to be very much an effect of imperfect teachers. The "I can not do math" self prejudice is not familiar to all cultures.

[^9]:    ${ }^{14}$ There was a famous misunderstanding when creationists understood that the theory of evolution is a theory in the sense of "it is possible" rather than a theory in the sense of "large body of knowledge".
    ${ }^{15}$ Of course, complicated mathematical objects will be described by complicated systems of sets.

[^10]:    ${ }^{16}$ The reason is simple: in order that $S^{\prime}$ and $Q$ be in the same position to one another as $S$ and $P$, we have to move $S$ to $S^{\prime}$ in the same way as we have moved $P$ to $Q$ !

[^11]:    ${ }^{17}$ Since $i d_{X}$ does not move any piece of $X$, the joint effect of $f$ and $i d_{X}$ is the same as effect of $f$ alone.

