1. (Very) Abridged historical context

1.1. **Gottfried Wilhelm Leibniz (1646–1716) and Isaac Newton (1642–1727).** They are credited with independently developing infinitesimal calculus. An explicit account was published by Leibniz starting with 1684. Newton used the techniques implicitly in several manuscripts around the same time and published an explicit account in 1704. The core ideas go back at least to Archimedes. At the heart of their understanding is the intuitive notion of *infinitesimal* (an infinitely small quantity, or a quantity smaller than any positive quantity but still non-zero) and *infinite* (an infinitely large quantity, or a quantity larger than any positive quantity). Understanding the notion of infinitesimal is problematic. No number value can be assigned to an infinitesimal; they are only endowed with magnitude (of smallness) which, again, cannot be measured in absolute terms but only relative to another infinitesimal. Hence, ratios of infinitesimals (such as derivatives) can have number values assigned to them and this is at the heart of differential calculus. On the other hand, multiplying an infinitesimal with an infinite can have a number value, as could have an infinite sum with infinitesimal terms (such as integrals) and this is at the heart of integral calculus. A sound theory of infinitesimals that does not appeal to intuition was not developed at that time. In fact, it requires a lot of sophistication to provide the theory of infinitesimals with a sound foundation and there is more than one way to do it (Abraham Robinson, 1960; Edward Nelson, 1977).

1.2. **Augustin-Louis Cauchy (1789–1857).** Through his lectures in the 1820s he starts developing a rigorous approach to infinitesimal calculus by replacing the use of infinitesimals with limiting arguments and, consequently, promoting an arithmetical treatment of the fundamental concepts of calculus. He defined the concept of limit through the $\varepsilon - \delta$ definition and he used limiting arguments to define rudimentary concepts of continuous, differentiable, and integrable function. For example, for Cauchy, continuity was a global concept (continuity on an interval) as opposed to local (continuity at a point). The word infinitesimal is used to refer to a variable quantity approaching zero (finite quantity that can be taken to be arbitrarily small). The concept of magnitude is still used but not specified. There were mistakes and false statements, some very significant, and in terms of clarity and rigor there was a lot room for improvement. Continuity was regarded as the most fundamental property but often extra hypotheses were tacitly used in the arguments and it was clear that, based on intuition, continuous functions were expected to have all the necessary properties.

*Bogdan Ion*, bion@pitt.edu
properties. Examples of properties intuitively used are (in today’s terminology) the intermediate value theorem and the Bolzano-Weierstrass theorem.

1.3. **Karl Weierstrass (1815–1897).** Through his lectures, starting in 1859-1860, he employed the \( \varepsilon, \delta, N \) approach to define all the fundamental concepts in calculus and applied these definitions systematically. In doing so he eliminated many of the obscurities present in Cauchy’s treatment. In particular, a clear distinction is made between uniform and pointwise continuity and statements such as the intermediate value theorem are rigorously justified. The concept of infinitesimal is completely phased out but the concept of real number and the identification of finite magnitudes with real numbers take its place. It was clear to Weierstrass that it was necessary to create a concept of real number (using rational numbers, for example) that captured the intuitive properties of Cauchy’s magnitudes. Weierstrass provided such a construction and proved that it had all the desired properties, such as the Bolzano-Weierstrass theorem.

1.4. **Richard Dedekind (1831–1916).** Out of concern that crucial properties of the real numbers depend on the intuitive understanding of \( \mathbb{R} \) as a geometrical line, in 1858 he obtained a construction of real numbers from rational numbers, employing what are nowadays called Dedekind cuts. The order relation and the arithmetical operations of real numbers arise naturally from this construction and the desired properties can be proved without any appeal to intuition. As rational numbers are obtained from integers, in 1872 Dedekind writes, in reference to the fundamental importance of the natural numbers, that every theorem of algebra and higher analysis could be rephrased as a theorem about the natural numbers and that indeed he had heard the great Dirichlet make this remark repeatedly. This raised the question of whether the natural numbers themselves can be obtained from something more fundamental. He proceeds to develop this idea and obtains in 1888 a construction of natural numbers using sets and logic. Giuseppe Peano published a similar account in 1889. The basic statements from which everything is developed are called the Dedekind-Peano axioms.

1.5. **Georg Cantor (1845-1918).** In 1871, Cantor and independently Heinrich Heine provided another construction of the real numbers as equivalence classes of Cauchy sequences of rational numbers. This work introduces for the first time the fundamental idea of completion as the formal version of the intuitive notion of filling gaps (in this case gaps between rational numbers).

While studying the problem of the uniqueness of the representation of a function as a trigonometric series, Cantor was led to consider infinite collections of points on which such a representation is not faithful. In the process he discovered that there are many kinds of infinite sets and their size account for different kinds of infinite numbers. This sits in contrast with the perception at the time that there is essentially one kind of infinite number that can be described as being larger than any finite number. In 1883-1885 he published his results and advocated for an extension of the arithmetical theory of the natural numbers that would include his infinite numbers (which turned out
to have counter-intuitive arithmetical properties). This naturally led to consider the notion of set as a primitive notion and a theory of sets as the foundation of mathematics. Setting up a coherent theory proved to be a formidable task as intuitive (in this context referred to as naive) notion of set leads very quickly to paradoxes, some of them known to Cantor.

1.6. **Ernst Zermelo (1871-1953).** The paradoxes of naive set theory called on the development of a firm axiomatic foundation of set theory. Zermelo provided a set of 7 axioms in 1908. By using them it was possible to obtain Dedekind’s development of the natural numbers, rational numbers, and real numbers, and, consequently, all the facts of calculus. Furthermore, it was possible to carry out Cantor’s study of sets, cardinals and ordinals.

Thoralf Skolem noticed that there are models of Zermelo’s set theory that do not allows large enough sets. He suggested in 1922 (also, independently by Adolf Fraenkel earlier in the same year) that a strengthening of one of the axioms was necessary. In 1925, John von Neumann added the axiom of regularity. The resulting set theory is called Zermelo-Fraenkel-Choice (ZFC). Skolem also pointed out that when founded in such an axiomatic way, set theory cannot remain a privileged logical theory; it is then placed on the same level as other axiomatic theories. Indeed, other axiomatic theories were soon constructed: John von Neumann developed a set theory with a finite number of first-order axioms which is based on functions as the primitive notion (as opposed to sets). The resulting set theory, after contributions of Bernays and Gödel is called von Neumann-Bernays-Gödel (NBG). Another axiomatic theory was developed by Whitehead and Russell in their book *Principia Mathematica* published during 1910-1913.

1.7. **David Hilbert (1862–1943).** Together with his student, Wilhelm Ackermann, he publishes in 1928 a book called *Principles of theoretical logic*. Turning to the questions associated with the axiomatic method they quickly recognize that the most important of the questions that arise are those of consistency, independence, and completeness of a system of axioms.

1.8. **Kurt Gödel (1906–1978).** Gödel’s 1930 paper proves the completeness of the axioms and rules of first-order logic, essentially as given in Hilbert and Ackermann. At the time it was well understood that all traditional mathematical proofs could be expressed in systems such as ZFC or Whitehead-Russell but their completeness (i.e. are they powerful enough to handle all future mathematics?) was still an open question. In his 1931 paper Gödel, using only elementary number theory, was able to construct true sentences (in either axiomatic system) about numbers which could not be proved using these axiomatic systems. Finally, he added a brief remark to the effect that if a sentence which expresses the consistency of such a system could be proved then the proof is not expressible within the system. This has widely been regarded as the end of Hilbert’s program to prove the consistency of mathematics by finitary means. In 1936, Gerhard Gentzen did succeed in proving the consistency of Dedekind-Peano arithmetic, but by using a non finitistic framework, namely transfinite induction.
1.9. The shift away from intuition to the new methodology based on precise definitions and clearly established inference rules brought not only fully reliable results but also true facts that were so counter-intuitive that generated shock and sometimes outrage in the mathematical community. Examples of such facts are the everywhere continuous, nowhere differentiable function (Weierstrass 1872; other examples were known as far back as the 1830’s, by Bolzano, for example), the space filling curve (Peano 1890; Hilbert 1891), the bijection between an interval and a \( n \)-dimensional cube (Cantor 1877).

2. Propositional logic

Natural language (e.g. English) is not precise enough for scientific purposes. Instead, some scientific fields (e.g. Mathematics) develop formal languages. These are artificial languages akin to computer programming languages, have a precise syntax, but also limited use. They often use elements of natural language (i.e. words) but always with a very well-determined meaning.

Example 2.1. The conjunction or, when used in natural language, typically depicts mutually exclusive alternatives: “You can have cake or ice cream”. In the formal language used for mathematics the previous statement has the meaning “You can have cake, or ice cream, or both”.

Definition 2.2. A logic is a formal language equipped with rules (called inference rules) for deducing the truth of a sentence from the truth of another sentence.

Next, we consider such a logic, called propositional logic.

Definition 2.3. A statement, also called a proposition, is a sentence that is either true (T) or false (F), but not both. If a sentence is a statement, we say that it has truth value (T or F, depending on the case).

Example 2.4. (Today is Friday), (We have tea at 3PM), (Lets drop this class), (Seven is prime), (Seven is even), (Steven is even), (This sentence is false).

Remark 2.5. The language of propositional logic contains words such as: ‘not’, ‘and’, ‘or’, ‘implies’, ‘equivalent’, ‘if and only if’. These have logical meaning that is similar, but not identical, with their natural language meaning. The logical meaning is precise and invariable. The natural language meaning often depends on the context. In mathematics, whenever such words appear in a statement it is assumed that they are used with their logical meaning. There are also symbols corresponding to each of them that can be alternatively used:

\[-\] for ‘not’ \quad \land \quad \text{for ‘and’}
\[\lor\] for ‘or’ \quad \Rightarrow \quad \text{for ‘implies’, ‘if...then...’} \quad \Leftrightarrow \quad \text{for ‘equivalent’, ‘if and only if’}

These words allow us that starting from some statements, called atoms or primitive statements, to construct more complicated statements, called formulas or compound statements.
**Definition 2.6.** If $A$ is a statement, $\neg A$ is the negation of $A$. If $A$ and $B$ are statements, $A \land B$ is the conjunction of $A$ and $B$.

**Example 2.7.** (Seven is not even), (Seven is prime) and (Seven is not even).

**Remark 2.8.** In logic $A \land B$ and $B \land A$ are always identical statements. In natural language this is not generally the case: (I ate at ‘Spice island’) and (I got sick) versus (I got sick) and (I ate at ‘Spice island’).

**Remark 2.9.** Syntax: formulas are obtained by repeatedly applying negation ($\neg A$) and conjunction ($A \land B$) and by using parentheses to specify the order of the operations. For example, $\neg(A \land B)$ and $\neg A \land B$ are different statements. More importantly, parentheses ensure that the resulting sentence is a statement. For example, $A \land B \lor C$ is not a statement, but $(A \land B) \lor C$ and $A \land (B \lor C)$ are statements.

We can tabulate the possible truth values of a formula in terms of the truth values of the atoms. The resulting table is called a truth table. Also, we can distinguish between two formulas by comparing their truth tables.

**Table 1.** Truth tables for $\neg A$, $A \land B$, and $\neg(\neg A \land \neg B)$

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$\neg A$</td>
<td>$A \land B$</td>
<td>$\neg(\neg A \land \neg B)$</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** Truth tables for $\neg A \land B$ and $\neg(A \land B)$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$B$</td>
<td>$\neg A$</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$B$</td>
<td>$A \land B$</td>
<td>$\neg(A \land B)$</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>

**Definition 2.10.** $A \lor B$ is defined as $\neg(\neg A \land \neg B)$. $A \Rightarrow B$ is defined as $\neg A \lor B$. $A \Leftrightarrow B$ is defined as $(A \Rightarrow B) \land (B \Rightarrow A)$.

**Definition 2.11.** A conditional statement is a statement of the form $A \Rightarrow B$. In mathematical writing, conditional statements appear in the form “If $A$ then $B$“. For such a conditional statement $A$ is said to be the hypothesis and $B$ the conclusion. Also we say that $B$ is a consequence of $A$, or we say that $B$ is a necessary condition for $A$, or we say that $A$ is a sufficient condition for $B$. 
\[
\begin{array}{c|cc}
A & B & (A \Rightarrow B) \land (A \Leftarrow B) \\
\hline
T & T & T \quad T \\
T & F & F \quad T \\
F & T & T \quad F \\
F & F & T \quad T \\
\end{array}
\]

Table 3. Truth table for \( A \iff B \)

**Remark 2.12.** In logic, when \( A \) is false, the statement \( A \Rightarrow B \) is true. This is not consistent with the use in natural language: the statement “(The Sun revolves around the Earth), therefore (lions are vegetarian).” would not be considered to be true.

**Definition 2.13.** Formulas that are true regardless of the truth values of their atoms is called a **tautology**. A rule of inference is a tautological (bi)conditional statement. Formulas that are false regardless of the truth values of their atoms is called a **contradiction**. A fallacy is a conditional statement that is not tautological. Two formulas are said to be logically equivalent if they have the same truth values.

**(Some) Rules of inference**

**Double negation:** \( \neg \neg A \iff A \)

**Simplification:** \( A \land B \Rightarrow A \)

**Addition:** \( A \Rightarrow A \lor B \)

**Disjunctive syllogism:** \( ((A \lor B) \land \neg A) \Rightarrow B \)

**Reductio ad absurdum:** \( ((\neg A \Rightarrow B) \land (\neg A \Rightarrow \neg B)) \iff A \)

**Modus ponens:** \( (A \land (A \Rightarrow B)) \Rightarrow B \)

**Modus tollens:** \( (\neg B \land (A \Rightarrow B)) \Rightarrow \neg A \)

**Contrapositive:** \( (A \Rightarrow B) \iff (\neg B \Rightarrow \neg A) \)

**Deduction principle:** \( ((A \land B) \Rightarrow C) \iff (\Rightarrow (B \Rightarrow C)) \)

**Hypothetical syllogism:** \( ((A \Rightarrow B) \land (B \Rightarrow C)) \Rightarrow (A \Rightarrow C) \)

**Case analysis:** \( ((A \lor B) \land (A \Rightarrow C) \land (B \Rightarrow C)) \Rightarrow C \)

**Distributivity:** \( (A \land (B \lor C)) \iff (A \land (B \lor C)) \land (A \land (B \lor C)) \iff (A \land B) \lor (A \land C) \)

**DeMorgan’s laws:** \( \neg (A \land B) \iff \neg A \land \neg B, \neg (A \lor B) \iff \neg A \land \neg B \)

**Some fallacies:** \( (A \land (A \lor B)) \Rightarrow \neg B, (B \land (A \Rightarrow B)) \Rightarrow A, (\neg A \land (A \Rightarrow B)) \Rightarrow \neg B \)

**Remark 2.14.** There are many tautological conditional statements (rules of inference). The list provided neither contains all of them nor the list is as short as possible. The list of independent rules
of inference is very short, the longer list is provided so that they available to be used in arguments (and they are used often, indeed) without having to deduce them every time.

**Definition 2.15.** A theorem is a statement of the form

\[
\text{If } H_1, \ldots, H_n \text{ then } C
\]

Each \( H_i \) is a statement called hypothesis and \( C \) is a statement called the conclusion.

**Definition 2.16.** A proof of the theorem

\[
\text{If } H_1, \ldots, H_n \text{ then } C
\]

is a finite sequence \( P_1, \ldots, P_N \) of statements with \( P_N = C \), and such that each of \( P_i \) is either a hypothesis, or the conclusion of a rule of inference which has as hypothesis one or more of the previous statements.

If the theorem \( \text{If } H_1, \ldots, H_n \text{ then } C \) is true we write

\[
H_1, \ldots, H_n \vdash C
\]

**Remark 2.17.** In current mathematical practice a proof is typically not entirely expressed in logic, but could use some natural language as long as every statement expressed in natural language can be potentially translated into first-order logic.

**Example 2.18 (Direct proof).** If (it is not sunny today) and (it is colder than yesterday) and (we will go swimming only if it is sunny) and (if we do not go swimming, then we will have a barbecue) and (if we will have a barbecue, then we will be home by sunset), then (we will be home by sunset).

For the proof, let us argue as follows. Consider the following propositions: \( A = \text{(it is not sunny today)} \), \( B = \text{(it is colder than yesterday)} \), \( C = \text{(we will go swimming)} \), \( D = \text{(we will have a barbecue)} \), \( E = \text{(we will be home by sunset)} \). Our statement is

\[
\text{If } A, B, (C \Rightarrow \neg A), (\neg C \Rightarrow D), (D \Rightarrow E) \text{ then } E
\]

We obtain the following sequence of statements

\[
\begin{align*}
A & \text{ (hypothesis)} \\
C \Rightarrow \neg A & \text{ (hypothesis)} \\
\neg C & \text{ (modus tollens)} \\
\neg C \Rightarrow D & \text{ (hypothesis)} \\
D & \text{ (modus ponens)} \\
D \Rightarrow E & \text{ (hypothesis)} \\
E & \text{ (modus ponens)}
\end{align*}
\]
which constitutes a proof of our statement. Therefore we may write

\[ A, B, (C \Rightarrow \neg A), (\neg C \Rightarrow D), (D \Rightarrow E) \vdash E \]

**Definition 2.19** (Types of proofs). Let \( H_1, \ldots, H_n \vdash C \) be a theorem. Depending on whether we prove this statement or an equivalent statement, we can distinguish among several types of proofs.

1. **Direct proof**: \( H_1, \ldots, H_n \vdash C \) is proved.
2. **Proof by contrapositive**: \( \neg C \vdash \neg (H_1 \land \ldots \land H_n) \) is proved.
3. **Proof by contradiction**: \( H_1, \ldots, H_n, \neg C \vdash D \) is proved, where \( D \) is a contradiction.

If the hypothesis is of the form \( (H_1, \ldots, H_n) \lor (K_1, \ldots, K_m) \) then we also have

4. **Proof by exhaustion**: \( H_1, \ldots, H_n \vdash C \) and \( K_1, \ldots, K_m \vdash C \) are proved.

As we progress through the material we will encounter many examples of either type of proof.

3. **First-order logic**

In mathematics we typically encounter more complicated sentences that depend on variables and whose truth value depends on the value of the variables. To handle such statements we need to consider an extension of propositional logic, called first-order logic.

**Definition 3.1.** A predicate, also called open statement, is a sentence with one or more variables such that when the variable is replaced by a value from a specified domain, also called universe of discourse, the resulting sentence is a statement in propositional logic (i.e. unambiguously true or false).

**Example 3.2.** Predicate: \( x \) is white. Domain: rabbits. Domain: cars. With the domain natural numbers, \((x \text{ is white})\) is not a predicate.

To be able to create statements in propositional logic from predicates we need to expand the lexicon of propositional logic by adding the following quantifiers:

\[ \exists \text{ for ‘for some’, ‘there exists... such that...’} \quad \forall \text{ for ‘for all’} \]

‘For some’ is called the existential quantifier, and ‘for all’ is called the universal quantifier.

**Example 3.3.** (Every child has a toy). Notation \( c, t \) variables, domain of \( c \): children, domain of \( t \): toys.

\[ \forall c (\exists t (t \text{ belongs to } c)) \]

**Remark 3.4.** Syntax: formulas are obtained by repeatedly applying negation (\( \neg A \)), conjunction (\( A \land B \)), and the existential quantifier (\( \exists x A(x) \)) and by using parentheses to specify the order of the operations. For example, \( \forall y A(y) \land \exists x B(x, y) \) is a valid formula and \( x(B \forall A) y \exists \) it is not.

We are now able to create sentences in first-order logic, by using statements in propositional logic, predicates, and quantifiers according to our rules of syntax.
Definition 3.5. \( \forall x A(x) \) is defined as \( \neg \exists x A(x) \).

Remark 3.6. \( \neg \forall x A(x) \iff \exists x \neg A(x), \neg \exists x A(x) \iff \forall x \neg A(x) \)

Remark 3.7. Note that the order in which the quantifiers appear is important. For example, if we switch the order of the quantifiers in Example 3.3 we obtain the statement 
\[
\exists t (\forall c (t \text{ belongs to } c))
\]
which in natural language reads (There is a toy that belongs to all children).

Definition 3.8. The free variables in a sentence are those variables that are not quantified. In first-order logic a statement, also called proposition, is a sentence having no free variables.

Definition 3.9. The truth domain of a predicate \( P(x) \) in the domain \( D \) is the collection of values of \( x \) from \( D \) for which \( P(x) \) is true. We use the notation 
\[
\{ x \in D \mid P(x) \}
\]
for the truth domain of \( P(x) \).

The formula \( (\exists x \in D)P(x) \) is true if and only if \( \{ x \in D \mid P(x) \} \) is not empty. The formula \( (\forall x \in D)P(x) \) is true if and only if \( \{ x \in D \mid P(x) \} = D \).

Example 3.10 (Negating statements in first-order logic). Consider variables \( c, t \) with domains: children, and toys, respectively. The following is a list of statements and their negation
\[
\begin{align*}
\forall c(\exists t(t \text{ belongs to } c)) & \quad \exists c(\forall t(t \text{ does not belong to } c)) \\
(\text{Every child has a toy}) & \quad (\text{There is a child with no toy}) \\
\exists t(\forall c(t \text{ belongs to } c)) & \quad \forall t(\exists c(t \text{ does not belong to } c)) \\
(\text{There is a toy that belongs to all children}) & \quad (\text{For any toy there is a child that does not own it}) \\
\forall c(\exists t(t \text{ belongs to } c) \Rightarrow (t \text{ is red})) & \quad \exists c(\forall t((t \text{ belongs to } c) \land (t \text{ is not red}))) \\
(\text{Every child has a red toy}) & \quad (\text{There is a child with no red toy})
\end{align*}
\]

4. Set theory

You will notice that whenever somebody attempts to explain what a set is, they will either try to give examples of sets, or appeal to your intuition, or use in the definition equivalent terms such as collection, group, family, bunch, etc. The explanation for this state of affairs is the fact that the notion of set is a primitive notion in mathematics. It means that virtually all concepts in mathematics are ultimately defined in terms of sets. It also means that sets are not to be defined in terms of other concepts.

Two extreme attitudes toward the subject are as follows. For one, the notion of set is left to our intuition. This leaves open the possibility that anything that can be specified in some fashion can be a set and this leads to paradoxes such as
**Russell's paradox:** Consider $B$ to be the set of all sets $x$ such that $x$ is not an element of itself. Then $B$ is an element of itself if and only if $B$ is not an element of itself.

This shows that not anything should be considered a set and prompts the warning that the notion of set should be “handled with care”, without going into further details. This approach is referred to as *naive set theory*. It is generally safe as it is unlikely to be naturally led into pathological considerations and constructions such as the one above.

At the other extreme, nothing is left to intuition. The most basic properties and constructions that are desirable are carefully spelled out as axioms and everything else is deduced from these using the inference rules of first order logic. This approach is referred to as *axiomatic set theory*. Despite being the proper treatment of the subject, the axiomatic set theory requires lengthy considerations which are usually reserved for specialized courses. Below, we will informally discuss the axioms that will play a role further on.

The most basic concept in set theory is that of *belonging*. It is a primitive concept (i.e. undefined). Sets have elements and when $x$ is an element of the set $S$ (i.e. $x$ belongs to $S$) we write $x \in S$. If some object $x$ is not an element of $S$ we write $x \not\in S$.

4.1. **Extension axiom.** *Two sets are equal if and only if they have the same elements.* This assures that sets are determined by their elements. If $A$ and $B$ are sets and every element of $A$ is also an element of $B$ we write $A \subset B$, or $A \subseteq B$. We say that $A$ is a subset of $B$. To indicate that $A$ is a subset of $B$ but not equal to $B$ we use the notation $A \subsetneq B$. The axiom of extension has the following extremely useful consequence.

**Theorem 4.1** (Double inclusion). Let $A$ and $B$ be two sets. Then $A = B$ if and only if $A \subset B$ and $B \subset A$.

4.2. **Selection axiom.** *For every set $A$ and every first-order logic formula $S(x)$ having $x$ with domain $A$ as its only free variable, the truth domain $B$ of $S(x)$ in $A$ is a set.* This assures that we can construct new sets in the fashion we intuitively do (by specifying properties of their elements) but it also rules out paradoxes such as Russell’s paradox. We denote $B$ as follows

$$B = \{ x \in A \mid S(x) \}$$

Note first that $B$ is a subset of $A$. In the intuitive approach to set theory, the ambiguity in what is allowed to be a subset of a set $A$ leads to contradictions. The selection axiom removes this ambiguity.

We can see the set $B$ that appears in Russell’s paradox is defined as follows

$$B = \{ x \text{ set} \mid x \not\in x \}$$
For this to fall under the provisions of the axiom there must be an ambient set $A$ for $x$. In this particular case $A$ should be set of all sets. If $A$ is a set, the paradox is the contradiction

$$B \in B \Leftrightarrow B \notin B$$

In the true conditional statement ($A$ is a set) $\Rightarrow (B \in B \Leftrightarrow B \notin B)$ the conclusion is false. Hence, the hypothesis is false. In other words $A$ is not a set. Therefore, the paradox is in fact a proof of the statement "the collection of all sets is not a set". We have our first non-example of set.

4.3. **Empty set axiom.** There exists a set with no elements. By the extension axiom, this set must be unique. We call it the empty set and we denote it by $\emptyset$. A statement equivalent to this axiom is *There exists a set*. since once a set exists, the selection axiom for $S(x)$ that is false for any $x$ implies the existence of the empty set. We also have an axiom that assures the existence of an infinite set so the empty set axiom is a consequence of that and it is only listed for practical reasons.

4.4. **Pair set axiom.** If $A$, $B$ are sets, there exists a set $\{A, B\}$ whose elements are $A$ and $B$.

4.5. **Union axiom.** If $I$ is a set and its elements $i \in I$ are themselves sets then there exists a set $\cup_{i \in I}i$ whose elements are the elements of the elements of $I$. In most circumstances this is applied in the following context. $I = \{A, B\}$, with $A, B$ sets. Then, their union is denoted $A \cup B$ and consists of the elements of $A$ and $B$.

4.6. **Power set axiom.** If $A$ is a set, there exists a set $P(A)$ whose elements are the subsets of $A$. For example, $P(\emptyset) = \{\emptyset\}$ since there is just one subset of $\emptyset$, $\emptyset$ itself. Note that $\emptyset$ has no elements but $\{\emptyset\}$ has precisely one element: $\emptyset$. Similarly, $P(P(\emptyset)) = \emptyset, \{\emptyset\}$ has two elements: $\emptyset$ and $\{\emptyset\}$.

4.7. **Infinite set axiom.** There is a set $A$ such that $\emptyset \in A$ and $(x \in A) \Rightarrow (\{x\} \in A)$. The smallest such set is

$$\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \ldots\}$$

4.8. **Other axioms.** The *Foundation axiom* rules out the existence of circular chains of sets (for example, no set is an element of itself) as well as infinitely descending chains of sets. The *Replacement axiom* postulates the existence of sets constructed as images of sets through correspondences that are expressible as first-order formulas. The last axiom is the *Choice axiom*. This is perhaps the most subtle of all axioms and large parts of mathematics depend on it. We postpone its discussion until it becomes unavoidable.

4.9. **Ordered pairs.** Define $(a, b) = \{\{a\}, \{a, b\}\}$. If $A, B$ are sets, define the Cartesian product of $A$ and $B$ as

$$A \times B = \{(a, b) \mid a \in A, b \in B\}$$

**Theorem 4.2.** If $(a, b) = (c, d)$ then $a = c$ and $b = d$. 
4.10. **Relations.** Let $A, B$ be sets. A relation between $A$ and $B$ is a subset $R \subset A \times B$. If $(a, b) \in R$ we write $aRb$. Let $R \subset A \times B$ and $S \subset B \times C$ be two relations between $A$ and $B$, and between $B$ and $C$, respectively. Define the relations $R^{-1} \subset B \times A$ and $S \circ R \subset A \times C$ as follows

$$ R^{-1} := \{(b, a) \mid aRb\}, \quad S \circ R := \{(a, c) \mid \exists b \in B, \ aRb \land bSc\} $$

Define also, the domain and, respectively the range of the relation $R$ as

$$ \text{Dom}(R) := \{a \in A \mid \exists b \in B, \ aRb\}, \quad \text{Range}(R) := \{b \in B \mid \exists a \in A, \ aRb\} $$

4.11. **Functions.** Set up the following notation

$$ \exists! x P(x) := \exists x P(x) \land (P(y) \land P(z) \Rightarrow y = z) $$

Let $A, B$ be sets and $R \subset A \times B$ a relation. We say that $R$ is a function if

1. $\text{Dom}(R) = A$
2. $\forall a \in A (\exists! b \in B \ (aRb))$