First Order Logic

November 17, 2016

1. The Notions of First-Order Logic

We first introduce the main concepts and notions of first order logic. First-order logic is intended to be the logic of "real mathematics." Propositional logic, while it models a form of mathematical reasoning, is not expressive enough to allow the mathematical reasoning used in general mathematics. In particular, the basic objects in propositional logic are propositional variables and truth values assignments. We would like our logic to be able to discuss more general mathematical objects or "structures." Also, in propositional logic we reason just with the binary connectives $\neg$, $\to$, whereas in general, mathematics reasoning with quantifiers is an important ingredient.

The overall plan, however, is still similar to that for propositional logic.

This is far from a complete survey, but rather a quick presentation of the central points. We first review the basic set-up of first-order logic, with an eye towards two cases of particular interest: set theory (which we have been doing all along) and number theory (or arithmetic) which will be of importance for the Gödel incompleteness theorem.

Definition 1.1. By a *language of first-order logic* we mean a collection $\mathcal{L} = \{R_i, f_i, c_i\}_{i \in \mathcal{I}}$ of relation symbols R_i, function symbols f_i, and constant symbols c_i. Usually we need only the countable case where $\mathcal{I} = \mathbb{N}$.

Each relation symbol R or function symbol f in a first-order language has an *arity* associated with it, with is a positive integer. The exact meaning of the arity is made clear below, but it is intending to denote the arity of a relation or function to which the symbol R or f corresponds.

Example 1. The language of set theory consists of a single binary relation symbol $\in$. So, $\mathcal{L} = \{\in\}$.

Example 2. The language of number theory consists of binary function symbols $+, \cdot, E$ (E is intended to denote the exponentiation function $(x, y) \to x^y$), a unary function symbol S (intending to denote the successor function $n \to n+1$), a binary relation symbol $<$, and a constant symbol $\mathbf{0}$. Thus, $\mathcal{L} = \{<, +, \cdot, E, S, \mathbf{0}\}$.

Note that the definition of a language is purely "syntactical," that is, no meaning is in anyway assigned to these symbols of the language (e.g., there is nothing that says $+$ somehow really represents what we think of a addition on the natural numbers).

For any first-order language $\mathcal{L}$, we consider also certain *logical symbols* which are always allowed in constructing formulas. These are variable symbols $\{x_i\}_{i \in \mathcal{I}}$ (frequently we need only the countable case $x_0, x_1, x_2, \dots$), connectives $\to$, $\neg$, a quantifier symbol $\forall$, parentheses $($, $)$, and a symbol $\approx$ for equality. We could (and do) choose to officially only use just the quantifier $\forall$, defining $\exists$ as $\neg\forall\neg$ (a more formal definition is given below).

We now define the formulas (or well-formed formulas, wffs) of the language $\mathcal{L}$. We first need to define the *terms* of the language. The terms of the language are the syntactical objects which are intended to denote actual objects (how they actually do this will be made clear below).

Definition 1.2. For $\mathcal{L}$ a first-order language, the *terms* of the language are defined recursively as follows:

1. Variable symbols x_i and constant symbols c_i are terms.
2. If $f \in \mathcal{L}$ is an n-ary function symbol and $t_1, \ldots, t_n$ are terms, then $f(t_1 \cdots t_n)$ is a term.

To improve readability, we usually write the term as $f(t_1, \ldots, t_n)$, that is we put commas in the list of subterms, but they are not officially part of the language.

Definition 1.3. The *formulas* (or wffs) of the language $\mathcal{L}$ are defined recursively as follows.

1. If $R \in \mathcal{L}$ is an n-ary relation symbol and $t_1, \ldots, t_n$ are terms, then $R(t_1 \cdots t_n)$ is a formula. Also, if t_1, t_2 are terms, then $\approx (t_1 t_2)$ is a formula. We often write this in the more readable form $t_1 \approx t_2$.
2. If ϕ, ψ are formulas then so are $(\phi \to \psi)$, $(\neg\phi)$.
3. If ϕ is a formula, then so is $(\forall x\, \phi)$, for any variable symbol x.

The formulas in case (1) are said to be the *atomic* formulas. Again, we ofter insert commas and write the atomic formula as $R(t_1, \ldots, t_n)$.

As in propositional logic, we must prove a unique readability theorem for the syntax.

Theorem 1.4. *(unique readability)*

(1) *We have unique readability for terms. That is, for every term u exactly one of the following three cases holds:*
   (a) *$u = c$ for some constant symbol in $\mathcal{L}$.*
   (b) *$u = x$ for some variable symbol in $\mathcal{L}$.*
   (c) *$u = f(t_1 \cdots t_n)$ for some n-ary function symbol f in $\mathcal{L}$, and terms $t_1, \ldots, t_n$.*

   In case (1c), the function symbol f and the subterms $t_1, \ldots, t_n$ are uniquely determined.

(2) *We have unique readability for wffs. That is, for every wff φ exactly one of the following cases hold:*
   (a) *$\varphi = (\neg\alpha)$ for some wff α.*
   (b) *$\varphi = (\alpha \to \beta)$ for some wffs α, β.*
   (c) *$\varphi = (\forall x \alpha)$ for some variable x and wff α.*

   In case (2a) the wff α is uniquely determined, in case (2b) the wffs α, β are uniquely determined, and in case (2c) the variable x and the wff α are uniquely determined.

Proof. Let K be the function defined on finite strings of symbols from the language $\mathcal{L}$ obtained by adding the values of K on the individual symbols, which are defined by: $K(c) = K(x) = 1$, $K(() = -1$, $K()) = 1$, $K(\neg) = 0$, $K(\to) = -1$, $K(f) = 1-n$ if f is an n-ary function symbol, $K(R) = 1 - n$ if R is an n-ary relation symbol, $K(\forall) = -1$.

We first prove unique readability for terms.

Claim 1. For every term t, $K(t) = 1$. If t' is a proper initial segment of t, then $K(t') < 1$.

Proof. The claim is proved by a straightforward induction on the terms. For the case $t = f(t_1 \cdots t_n)$, we have $K(t) = K(f) + K(() + K(t_1) + \cdots + K(t_n) + K()) = (1-$

$n)+(-1)+n+1 = 1$. If t' is a proper initial segment of t of the form $t' = f(t_1 \cdots t_j t'_j$, then $K(t') = (1-n) + (-1) + j + K(t'_j) < (1-n) + (-1) + j + 1$ by induction. Since $j \leqslant n-1$, $K(t') < 0$. If $t = f(t_1 \cdots t_n$, then $K(t) = (1-n) + (-1) + n = 0$. The other cases are similar. □

Clearly the case $t = c$, $t = x$, $t = f(t_1 \cdots t_n)$ are mutually exclusive, and in the first two cases the c and x are uniquely determined by t. Suppose t is a term and $t = f(t_1 \cdots t_n) = g(u_1 \cdots u_m)$ for some n-ary function symbol f, m-ary function symbol g, and terms $t_1, \dots, t_n, u_1, \dots, u_m$. Since t begins with the symbol f and also the symbol g, we have $f = g$. Thus, $n = m$ as well. So, $t = f(t_1 \cdots t_n) = f(u_1 \cdots u_n)$. Thus, $t_1 \cdots t_n = u_1 \cdots u_n$. t_1 is the least initial t' segment of t for which $K(t') = 1$, and likewise for u_1. Thus, $t_1 = u_1$. So, $t_2 \cdots t_n = u_2 \cdots u_n$. Continuing in this manner we get that $t_i = u_i$ for all $i = 1, \dots, n$.

Next we show unique readability for wffs.

Claim 2. For every wff φ, $K(\varphi) = 1$. If φ' is a proper initial segment of φ, then $K(\varphi') < 1$.

Proof. The claim is proved by a straightforward induction on the wffs. In the case $\varphi = (\forall x\, \alpha)$, $K(\varphi) = K(() + K(\forall) + K(x) + K(\alpha) + K())= (-1) + (-1) + 1 + K(\alpha) + 1 = K(\alpha) = 1$ by induction. The remaining cases are similar, and left as an exercise. □

It is clear form the definition that no wff begins with a $\neg$ or $\forall$ (they all must begin with a left parenthesis). Form this it follows that the three cases in the definition of wff are mutually exclusive. For example, suppose $\varphi = (\alpha \to \beta) = (\forall x\, \gamma)$. Then $\alpha \to \beta$ is equal to the string $\forall x\, \gamma$. This says that α begins with a $\forall$, a contradiction. The other cases are similar. If $(\alpha \to \beta) = (\gamma \to \delta)$, then $\alpha \to \beta$ is equal to $\gamma \to \delta$. Then α is the least initial segment u of the first string for which $K(u) = 1$, and likewise for γ. So, $\alpha = \gamma$, and it follows that $\beta = \delta$. The other cases are easier. □

Exercise 1. Complete the proof of Claim 2 for the cases $\varphi = (\neg\alpha)$, $\varphi = (\alpha \to \beta)$.

Exercise 2. Show that we can eliminate parentheses entirely in first-order logic by using reverse Polish notation. Define terms exactly as before. For the wffs, take as definitions: $\varphi = \neg\alpha$, $\varphi = \to \alpha\beta$, and $\varphi = \forall x \alpha$. Use the same K function.

We use the usual abbreviations $\wedge$, $\vee$ as we did in propositional logic. Again, they are not officially part of the language, but we will write $(\alpha \wedge \beta)$ and $(\alpha \vee \beta)$ as if they were.

An occurrence of a variable x in a formula ϕ is said to be *free* if it is not within the scope of a $\forall x$ or $\exists x$ quantifier. More precisely:

Definition 1.5. We define by recursion of the formulas:

(1) Any occurrence of x in an atomic formulas is free.
(2) x occurs free in $(\phi \wedge \psi)$ where it occurs free in ϕ and where it occurs free in ψ. x occurs free in $(\neg\phi)$ where it occurs free in ϕ.
(3) x occurs free in $(\forall y\, \phi)$ where it occurs free in ϕ if $y \neq x$. If $y = x$, then x does not occur free in $(\forall y\, \phi)$.

We say x occurs free in ϕ if it has a free occurrence in ϕ. We say a *sentence* is a formulas with no free variables.

We usually write $\phi(x_1, \ldots, x_k)$ to denote a formula ϕ whose free variables are among $x_1, \ldots, x_k$.

Intuitively, a formulas $\phi(x_1, \ldots, x_k)$ is intended to an assertion about the objects represented by $x_1, \ldots, x_n$. In particular, a sentence is intending to be a statement whose truth can be ascertained without any knowledge of what certain variables represent. We have not yet officially assigned any meaning to formulas yet, they are currently purely syntactical.

Example 3. If $\mathcal{L}$ is the language of set theory, the only terms are the variables x_i. The only atomic formulas are $x_i \approx x_j$ and $x_i \in x_j$. The axioms of set theory, ZF or ZFC, will all be sentences in this language. Since set theory will incorporate all of traditional mathematics, all statements and theorems of traditional mathematics can be expessed as sentences in this language.

Example 4. Suppose now $\mathcal{L}$ is the language of number theory. This language is intended to make statements about the natural numbers $\mathbb{N}$ (which hasn't been give a rigorous axiomatization yet). The formula

$$\phi = (\neg(x \approx 0) \wedge \neg(x \approx S(0)) \wedge \forall m\ \forall n\ (x \approx m \cdot n \to (m \approx S(0) \vee n \approx S(0))))$$

is a formula with free variable x which attempts to assert that x is prime. The sentence

$$\psi = \forall m\ \exists n\ \exists k\ (n > m \wedge \phi(n) \wedge \phi(k) \wedge k \approx S(S(n)))$$

asserts the twin-prime conjecture.

Exercise 3. Write down formulas or sentences which do the following.

(1) A formula $\varphi(x)$ in the language of number theory which asserts that x is the sum of three squares, and a sentence ψ which asserts that every natural number is the sum of four squares.

(2) A formula $\varphi(x)$ in the language of set theory which asserts that every element of an element of x is an element of x (this is called x being *transitive*).

(3) A formula $\varphi(x)$ in the language $\mathcal{L} = \{+, \cdot\}$ which, if we interpret $+$ and $\cdot$ as the usual addition and multiplication in $\mathbb{R}$, holds iff $x > 0$.

In order to assign meaning to a formulas, we must have a "universe" of objects in which to interpret the quantifiers, interpretations of the relation, function, and constant symbols of the language, and if the formula has free variables we also have to know how interpret the free variables. This is made precise in the following definition of structure.

Definition 1.6. A *structure* for a first-order language $\mathcal{L} = \{R_i, f_i, c_i\}_{i \in \omega}$ is an object of the form $\mathfrak{A} = (A; R_i^A, f_i^A, c_i^a)$ where A is a non-empty set, R_i^A is an n-ary relation on A (where n is the arity of the relation symbol R_i), $f_i^A \colon A^n \to A$ is an n-ary function (again, n is the arity of f_i), and $c_i^A \in A$.

Thus, a structure provides a universe A as well an an interpretation of the symbols of the language $\mathcal{L}$. Structures may be thought of as fairly general mathematical objects. For example, consider the language of group theory, which consists of a single binary function symbol $\cdot$. A structure for the language is a set A with a binary operation $(a, b) \to a \cdot b \in A$ on A. A group, for example, would be such a structure (but of course not all structures would be groups).

As another example, for $\mathcal{L}$ the language of number theory, the "usual" natural numbers would be a structure $\mathbb{N} = \{\mathbb{N}; +^{\mathbb{N}}, \cdot^{\mathbb{N}}, E^{\mathbb{N}}, <^{\mathbb{N}}, S^{\mathbb{N}}, 0^{\mathbb{N}}\}$ for the language, where $+^{\mathbb{N}}$ denotes the usual addition on $\mathbb{N}$, etc. Of course, there are many other structures for this language. [note: what we mean by the "usual" natural numbers is vague and undefined here. We can think of this as meaning the finite ordinals of a model of ZFC set theory which we suppress mentioning and identify with the metatheory].

A structure will be sufficient to decide the truth or falsity of a sentence, but for a formula $\phi(x_1, \ldots, x_n)$ with free variable, we need also a map $s\colon \text{var} \to A$ interpreting the variables (or at least the free variables in ϕ). We wish to define precisely the meaning of " the structure $\mathfrak{A}$ satisfies ϕ at the variable assignment s," which we will denote as $\mathfrak{A} \models \phi[s]$. The formal definition follows.

Definition 1.7. Let $\mathfrak{A}$ be a structure for the first-order language $\mathcal{L}$, ϕ a formula, and $s\colon \text{Var} \to |\mathfrak{A}|$. We define $\mathfrak{A} \models \phi[s]$ follows.

First we extend the interpretation function s to terms t.

(1) If $x \in \text{var}$, then $s(x)$ is already defined. If c is a constant symbol, then $s(c) = c^{\mathfrak{A}}$.
(2) If $u = f(t_1, \ldots, t_n)$ is a term, then $s(u) = f^{\mathfrak{A}}(s(t_1), \ldots, s(t_n))$.

Next we define the notion $\mathfrak{A} \models \phi[s]$ by recursion on ϕ as follows.

(1) If ϕ is the atomic formula $R(t_1, \ldots, t_n)$, then $\mathfrak{A} \models \phi[s]$ iff $R^{\mathfrak{A}}(s(t_1), \ldots, s(t_n))$. If ϕ is the atomic formula $t_1 \approx t_2$, then $\mathfrak{A} \models \phi[s]$ iff $s(t_1) = s(t_2)$.
(2) $\mathfrak{A} \models (\phi \to \psi)[s]$ iff $\mathfrak{A} \not\models \phi[s]$ or $\mathfrak{A} \models \psi[s]$. $\mathfrak{A} \models (\neg\phi)[s]$ iff it is not the case that $\mathfrak{A} \models \phi[s]$.
(3) $\mathfrak{A} \models \forall x \phi[s]$ iff for every $a \in A$ we have that $\mathfrak{A} \models \phi[s(x|a)]$, where $s(x|a)(y) = s(y)$ if $y \neq x$ and $s(x|a)(x) = a$.

The above definition contains no surprises; it simply formalizes the usual notion of satisfaction of a statement in a mathematical structure.

If Γ is a collection of formulas, we write $\mathfrak{A} \models \Gamma[s]$ to mean $\mathfrak{A} \models \phi[s]$ for all $\phi \in \Gamma$.

Fact 1.8. Let $\mathcal{L}$ be a language of first-order logic, and $\mathfrak{A}$ a structure for $\mathcal{L}$. If $s_1, s_2\colon \text{Var} \to |\mathfrak{A}|$ agree at all the free variables of φ, then $\mathfrak{A} \models \varphi[s_1]$ iff $\mathfrak{A} \models \varphi[s_2]$

Proof. We first prove by induction on the terms that if s_1, s_2 agree at all the variables in a term t, then $s_1(t) = s_2(t)$. This is a straightforward induction. The inductive step is a s follows. Suppose $t = f(t_1, \cdots, t_n)$. All of the variables in a t_i occur also in t, so s_1, s_2 agree on all the variables of t_i. By induction, $s_1(t_i) = s_2(t_i)$ for all subterms t_i. But then $s_1(t) = f^{\mathfrak{A}}(s_1(t_1), \cdots, s_1(t_n)) = f^{\mathfrak{A}}(s_2(t_1), \cdots, s_2(t_n)) = s_2(t)$.

We now prove the fact by induction on the wff φ. Suppose first φ is atomic, say $\varphi = R(t_1 \cdots t_n)$. By definition, all of the variables in any subterm t_i occur free in φ. So, s_1, s_2 argree on all the variables in φ. By the above result for terms, $s_1(t_i) = s_2(t_i)$ for $1 \leqslant i \leqslant n$. But then $\mathfrak{A} \models \varphi[s_1]$ iff $R^{\mathfrak{A}}(s_1(t_1), \ldots, s_1(t_n))$ iff $R^{\mathfrak{A}}(s_2(t_1), \ldots, s_2(t_n))$ iff $\mathfrak{A} \models \varphi[s_2]$.

We consider one of the boolean connective cases, say $\varphi = (\alpha \to \beta)$. The variables free in φ are the variables free in α together with the variables free in β. Thus, s_1, s_2 agree on all the free variables of α and β. By induction, $\mathfrak{A} \models \alpha[s_1]$ iff $\mathfrak{A} \models \alpha[s_2]$, and likewise for β. So, $\mathfrak{A} \models \varphi[s_1]$ iff ($\mathfrak{A} \not\models \alpha[s_1]$ or $\mathfrak{A} \models \beta[s_1]$) iff ($\mathfrak{A} \not\models \alpha[s_2]$ or $\mathfrak{A} \models \beta[s_2]$) iff $\mathfrak{A} \models \varphi[s_2]$.

Consider finally the case $\varphi = \forall x\, \psi$. All the free variables of ψ occur free in φ, except perhaps x (if it is free in ψ). Thus, s_1 and s_2 agree at the free variables of ψ except at x (if it is free in ψ). Now, $\mathfrak{A} \models \varphi[s_1]$ iff for all $a \in |\mathfrak{A}|$ we have $\mathfrak{A} \models \psi[s_1(x|a)]$. Note that $s_1(x|a)$ and $s_2(x|a)$ agree at x, as well as the other free variables of ψ. Thus, $s_1(x|a)$ and $s_2(x|a)$ agree at all the free variables of φ. By induction, $\mathfrak{A} \models \psi[s_1(x|a)]$ iff $\mathfrak{A} \models \psi[s_2(x|a)]$ for all $a \in |\mathfrak{A}|$. So, $\mathfrak{A} \models \varphi[s_1]$ iff for all $a \in \mathfrak{A}$ $\mathfrak{A} \models \psi[s_1(x|a)]$ iff for all $a \in \mathfrak{A}$ $\mathfrak{A} \models \psi[s_2(x|a)]$ iff $\mathfrak{A} \models \varphi[s_2]$. □

We now introduce the notion of logical implication.

Definition 1.9. Let Γ be a collection of formulas in a first-order language $\mathcal{L}$, and let ϕ be a formula. We write $\Gamma \models \phi$ if for every structure $\mathfrak{A}$ for $\mathcal{L}$ and every $s\colon \text{var} \to |\mathfrak{A}|$, if $\mathfrak{A} \models \Gamma[s]$, then $\mathfrak{A} \models \phi[s]$.

Thus, $\Gamma \models \phi$ if every structure satisfying Γ also satisfies ϕ. Frequently we think of Γ as being the axioms for some theory we are studying.

Example 5. For $\mathcal{L}$ the language of group theory (i.e., a single binary function symbol, but for convenience we now add a constant symbol e to the language as well), let Γ be the following set of sentences:

$\forall x\ \forall y\ \forall z\ x \cdot (y \cdot z) \approx (x \cdot y) \cdot z$
$\forall x\ (x \cdot e \approx x)$
$\forall x\ \exists y (x \cdot y \approx e)$.

Thus, Γ is the usual set of axioms for groups. If ϕ is a sentence in the language of group theory, then $\Gamma \models \phi$ iff ϕ is true in all groups.

Exercise 4. Write down a set Γ of wffs in the empty language $\mathcal{L}$ such that $\mathfrak{A} \models \Gamma$ iff $|\mathfrak{A}|$ is infinite. We will see below that there is no Γ, in any language, such that $\mathfrak{A} \models \Gamma$ iff $|\mathfrak{A}|$ is finite.

Example 6. Let $\mathcal{L} = \{<\}$ be the language with one binary relation symbol $<$. Let Γ be the following set of wffs:

(1) $\forall x\, \neg(x < x)$ (irreflexive)
(2) $\forall x\, \forall y\, \forall z\, ((x < y \wedge y < z) \to x < z)$ (transitive)
(3) $\forall x\, \forall y\, ((x < y) \vee (y < x) \vee (x \approx y))$ (connected)

The wffs in Γ assert that $<$ is strict linear order. There are models of Γ which are both finite and infinite, for example the set $\{0, 1, 2, \dots, n\}$ with the usual ordering, and $\mathbb{N}$ with the usual ordering. Consider $\Gamma' = \Gamma \cup \{\forall x\, \exists y\, (x < y)\}$. The extra wff asserts there is no largest element. Any model of Γ' must be infinite.

Exercise 5. Write down a set of wffs Γ in the language $\mathcal{L} = \{R\}$, where R is a binary relation symbol, such that $\mathfrak{A} \models \Gamma$ iff $\mathfrak{A}$ is a graph with no cycles. Do you think it is possible to make Γ finite?

If $\varphi(x_1, \dots, x_n)$ is a formula with free variables among $x_1, \dots, x_n$, $\mathfrak{A}$ is a structure, and $a_1, \dots, a_n \in A = |\mathfrak{A}|$, we write $\mathfrak{A} \models \varphi(a_1, \dots, a_n)$ to abbreviate $\mathfrak{A} \models \varphi[s]$ where $s\colon \text{Var} \to A$ is such that $s(x_i) = a_i$ for $1 \leqslant i \leqslant n$.

Definition 1.10. If $\mathfrak{A}$ is a structure, say a set $B \subseteq A^k = |\mathfrak{A}|^k$ is *definable* (from parameters) if there is a formula $\varphi(x_1, \dots, x_k, y_1, \dots, y_\ell)$ and $a_1, \dots, a_\ell \in A$ such that $(b_1, \dots, b_k) \in B$ iff $\mathfrak{A} \models \varphi(b_1, \dots, b_k, a_1, \dots, a_\ell)$.

Example 7. Let $\mathfrak{N} = (\omega; +^{\mathbb{N}}, \cdot^{\mathbb{N}})$ be the structure consisting on the natural numbers (denoted ω here) with the usual operations of addition and multiplication. Then $0^{\mathfrak{N}}$, $S^{\mathfrak{N}}$, $<^{\mathfrak{N}}$, $E^{\mathfrak{N}}$ are all definable (without parameters) in $\mathbb{N}$. We can define

0 by the formula $\varphi(x) = \forall y\,(x + y \approx y)$. We can define $<$ by $\rho(x, y) = \exists z\,(z \not\approx 0 \wedge x + z \approx y)$. We can define 1 by $\psi(x) = \forall y\,(x \cdot y \approx y)$. It follows that we can define any singleton set $A = \{k\}$, and in fact any finite or cofinite set. The set of primes is definable. Note that the collection of definable sets is closed under complements, finite unions, and finite intersections. The fact that E is definable takes more work, and will be shown later. We will also show that every *computable* function or relation is definable. Note that there are only countably many definable functions and relations, while the collection of all functions/relations on ω has size $c = 2^{\omega}$.

Example 8. Let $\mathfrak{R} = (\mathbb{R}; +^{\mathbb{R}}, \cdot^{\mathbb{R}})$ be the set of reals with the usual operations of addition and multiplication. $\{0\}$ is clearly definable, being the unique additive identity. The set $P \subseteq \mathbb{R}$ of positive real numbers is definable by $\varphi(x) = \exists y\,(y \cdot y \approx x) \wedge x \not\approx 0$. It follows that the usual linear ordering $<$ on $\mathbb{R}$ is definable in $\mathfrak{R}$.

Exercise 6. Let $\mathfrak{Z} = (\mathbb{Z}; +^{\mathbb{Z}}, \cdot^{\mathbb{Z}})$ be the integers with the usual operations of addition and multiplication. Show that $<$ and the set $A = \mathbb{N} = \{a \in \mathbb{Z} : a \geqslant 0\}$ are definable without parameters in $\mathfrak{A}$ (hint: use Lagrange's theorem in number theory).

1.1. **Homomorphisms.** The algebraic notion of homomorphism can be given in this general context.

Definition 1.11. Let $\mathfrak{A}$, $\mathfrak{B}$ be two structures for a first-order language $\mathcal{L} = \{R_i, f_i, c_i\}$. A *homomorphism* π from $\mathfrak{A}$ to $\mathfrak{B}$ is a map $\pi\colon A \to B$ satisfying: $R^{\mathfrak{A}}(a_1, \dots a_n)$ iff $R^{\mathfrak{B}}(\pi(a_1), \dots \pi(a_n))$, $\pi(f^{\mathfrak{A}}(a_1, \dots, a_n)) = f^{\mathfrak{B}}(\pi(a_1), \dots, \pi(a_n))$, and $\pi(c^{\mathfrak{A}}) = c^{\mathfrak{B}}$ for all $a_1, \dots, a_n \in A$ and all relation, function, and constant symbols R, f, c in $\mathcal{L}$. An *isomorphism* between $\mathfrak{A}$ and $\mathfrak{B}$ is a one-to-one, onto, homomorphism. An *automorphism* of $\mathfrak{A}$ is an isomorphism from $\mathfrak{A}$ to itself.

We can use homomorphisms and automorphisms of structures to show certain sets and functions are not definable in a given structure. The method uses the following basic result.

Theorem 1.12. *Let $\mathfrak{A}$, $\mathfrak{B}$ be structures for a first order language $\mathcal{L}$. Let $\varphi(x_1, \dots, x_n)$ be a wff of $\mathcal{L}$, with free variables among $\{x_1, \dots, x_n\}$. If $\pi\colon \mathfrak{A} \to \mathfrak{B}$ is an isomorphism (i.e., a homomorphism which is one-to-one and onto), then for any $a_1, \dots, a_n \in |\mathfrak{A}|$ we have $\mathfrak{A} \models \varphi(a_1, \dots, a_n)$ iff $\mathfrak{B} \models \varphi(\pi(a_1), \dots, \pi(a_n))$.*

Proof. For any $s\colon \mathrm{Var} \to |\mathfrak{A}|$, let $\pi(s)\colon \mathrm{Var} \to |\mathfrak{B}|$ be defined by $\pi(s)(x) = \pi(s(x))$.

First prove by induction on the terms that for any s we have that $\pi(s(t)) = (\pi(s))(t)$. For variables this is true by definition of $\pi(s)$. For constant symbols, $\pi(s(c)) = \pi(c^{\mathfrak{A}}) = c^{\mathfrak{B}} = (\pi(s))(c)$. For the inductive step, say $t = f(t_1, \dots, t_n)$. Then

$$\begin{aligned}
\pi(s(t)) &= \pi(f^{\mathfrak{A}}(s(t_1), \dots, s(t_n)) \\
&= f^{\mathfrak{B}}(\pi(s(t_1)), \dots, \pi(s(t_n))) \\
&= f^{\mathfrak{B}}((\pi(s))(t_1), \dots, (\pi(s))(t_n)) \\
&= (\pi(s))(t).
\end{aligned}$$

Next we prove the theorem by induction on φ. Suppose first that φ is atomic. Say $\varphi = R(t_1, \dots, t_n)$. Then $\mathfrak{A} \models \varphi[s]$ iff $R^{\mathfrak{A}}(s(t_1), \dots, s(t_n))$ iff $R^{\mathfrak{B}}(\pi(s(t_1), \dots, \pi(s(t_n))$

(as π is a homomorphism) iff $R^{\mathfrak{B}}((\pi(s))(t_1), \ldots, (\pi(s))(t_n))$ (by the result for terms) iff $\mathfrak{B} \models \varphi[\pi(s)]$. If $\varphi = t \approx u$, then $\mathfrak{A} \models \varphi[s]$ iff $S(t) = s(u)$ iff $\pi(s(t)) = \pi(s(u))$ (as π is one-to-one) iff $(\pi(s))(t) = (\pi(s))(u)$ (by the result for terms) iff $\mathfrak{B} \models \varphi[\pi(s)]$.

The boolean cases follow immediately by induction.

Suppose $\varphi = \forall x\, \psi$. Then $\mathfrak{A} \models \varphi[s]$ iff for all $a \in |\mathfrak{A}|$ we have $\mathfrak{A} \models \psi[s(x|a)]$. By induction, for a given a this happens iff $\mathfrak{B} \models \psi[\pi(s(x|a))]$. Now, $\pi(s(x|a)) = \pi(s)(x|\pi(a))$. Now, for all $a \in \mathfrak{A}$ $\mathfrak{B} \models \psi[\pi(s)(x|\pi(a))]$ iff for all $b \in \mathfrak{B}$ $\mathfrak{B} \models \psi[\pi(s)(x|b)]$ as π is onto. The last statement is the definition of $\mathfrak{B} \models \varphi[\pi(s)]$.

□

Consider the structure $\mathfrak{N}_m = (\mathbb{N}; \cdot)$. Note that 0 and $1 = S(0)$ are definable in this structure. Let P be the set of primes. Let $\pi \colon P \to P$ be any bijection. Let π' be the extension of π to $\mathbb{N}$ defined multiplicitively, that is, $\pi'(p_1^{a_1} \cdots p_k^{a_k}) = \pi'(p_1)^{a_1} \cdots \pi'(p_k)^{a_k}$. Easily π is an isomorphism from $\mathfrak{N}_m$ to itself (where we define $\pi'(0) = 0$). Note that P is definable in $\mathfrak{N}_m$, so every automorphism of $\mathfrak{A}_m$ must send P to P. Since a homomorphism must preserve multiplication, we see in fact that every automorphism of $\mathfrak{A}_m$ is of the form π' for some permutation π of P.

If now follows that $+$ is not definable in $\mathfrak{N}_m$, since, for example, $1 + 1 = 2$, but $\pi(1) + \pi(1) \neq \pi(2)$ if $\pi(2) = 3$, since $\pi(1) = 1$. We also have that $\{2\}$ is not definable, since there is a π which sends 2 to 3. Likewise it follows that the ordering $<$ is not definable in $\mathfrak{A}_m$. Thus we have:

Theorem 1.13. *Let $\mathfrak{N}_m = (\mathbb{N}; \cdot)$. Then $+_{\mathbb{N}}$ and $<_{\mathbb{N}}$ are not definable in $\mathfrak{N}_m$.*

Remark 1.14. It is also true that $\cdot_{\mathbb{N}}$ is not definable in the structure $\mathfrak{N}_a = (\mathbb{N}; +_{\mathbb{N}})$, but this cannot be shown using homomorphisms, as the only automorphism of the structure $\mathfrak{N}_a$ is the identity, as each $\{k\}$ is definable in $\mathfrak{N}_a$.

Exercise 7. Let $a_1, \ldots, a_k$ be positive integers. Let $A_{\vec{a}}$ be the set of positive integers n such that $n = p_1^{a_1} \cdots p_k^{a_k}$ for some distinct primes $p_1, \ldots, p_k$ (the p_j are not necessarily increasing). Show that $A_{\vec{a}}$ is definable in $\mathfrak{A}_m$.

Exercise 8. Show that the only automorphism of the structure $\mathfrak{R} = (\mathbb{R}; +_{\mathbb{R}}, \cdot_R)$ is the identity. [hint: If π is an automorphism of $\mathfrak{R}$, show that $\pi(q) = q$ for all $q \in \mathbb{Q}$, and use this along with the fact that $<$ is definable in $\mathfrak{R}$.] The argument given below, on the other hand, will show that there are many automorphisms of the complex numbers $\mathfrak{C} = (\mathbb{C}; +, \cdot)$.

From Example 8 we have that the usual linear ordering $<$ on $\mathbb{R}$ is definable over $\mathfrak{R}$. We show that there is no definable linear order over $\mathfrak{C} = (\mathbb{C}; +, \cdot)$, the complex numbers with the usual operations of addition and multiplication. Let $T \subseteq \mathbb{C}$ be a transcendence base for $\mathbb{C}$ over $\mathbb{Q}$, that is, T is a maximal set of algebraically independent elements (we use AC to get this). Let $\pi \colon T \to T$ be a permutation of T. Since $\mathbb{C}$ is algebraically closed, π extends to an automorphism $\pi \colon \mathbb{C} \to \mathbb{C}$ by a standard argument from field theory [build $\pi = \bigcup_{\alpha < 2^\omega} \pi_\alpha$. At even stages α, pick the least (in some wellorder of $\mathbb{C}$) $z_\alpha \in \mathbb{C}$ not in $\mathbb{C}_{<\alpha} = \bigcup_{\beta<\alpha} \operatorname{dom}(\pi_\beta)$ and let p_α be its minimal polynomial over $\mathbb{C}_{<\alpha}$. Let $\mathbb{C}_\alpha = \mathbb{C}_{<\alpha}(z_\alpha)$, and define $\pi_\alpha(z_\alpha)$ to be some root of $\pi(p)$ over $\mathbb{D}_{<\alpha} = \bigcup_{\beta<\alpha} \pi_\beta(C_\beta)$. Each π_α extends to a field isomorphism $\pi_\alpha \colon \mathbb{C}_\alpha \to \mathbb{D}_\alpha = D_{<\alpha}(\pi_\alpha(z))$. At odd stages, pick $z \in \mathbb{C} - \mathbb{D}_{<\alpha}$ and proceed similarly.]

Now, if $<$ were a definable over $\mathfrak{C}$ linear ordering of $\mathbb{C}$, then we would have $t_1 < t_2$ iff $\pi'(t_1) < \pi'(t_2)$ for any $t_1, t_2 \in T$ and permutation $\pi \colon T \to T$. But we can take π to switch t_1 and t_2, so this is a contradiction. We summarize this in the next theorem.

Theorem 1.15. *Let $\mathfrak{R} = (\mathbb{R}; +, \cdot)$, and $\mathfrak{C} = (\mathbb{C}; +, \cdot)$ be the usual structures for the real and complex numbers. Then the usual linear ordering of $\mathbb{R}$ is definable in $\mathfrak{R}$, but there is no linear ordering of $\mathbb{C}$ definable in $\mathfrak{C}$.*

The next theorem also illustrates the use of homomorphisms.

Theorem 1.16. *Multiplication on the reals, $\cdot_{\mathbb{R}}$, is not definable in the structure $\mathfrak{R}_a = (\mathbb{R}; +)$. Addition on the reals, $+_{\mathbb{R}}$, is not definable in the structure $\mathfrak{R}_m = (\mathbb{R}; \cdot)$. A similar statement holds for the complex numbers.*

Proof. We consider the case of the reals, the complex case being similar. Let $B \subseteq \mathbb{R}$ be a basis for $\mathbb{R}$ over $\mathbb{Q}$ (a *Hamel basis*, we use AC to get this). Recall this means B is a maximal set of linearly independent elements over $\mathbb{Q}$. Let $\pi\colon B \to B$ be a bijection. Then π extends to an automorphism $\pi'\colon \mathfrak{R}_a \to \mathfrak{R}_a$. Namely, if $x \in \mathbb{R}$, then there is a unique representation $x = q_1 z_1 + \cdots + q_n z_n$, where $z_i \in B$. Set $\pi(x) = q_1\pi(z_1) + \cdots + q_n\pi(z_n)$. This is easily welldefined and an automorphism of $\mathfrak{R}_a$. Suppose $\cdot_{\mathbb{R}}$ were definable over $\mathfrak{R}_a$. Let $t_1, t_2 \in B$ be algebraically independent elements (which easily exist as B has size 2^ω). Let $t_1 \cdot t_2 = q_1 z_1 + \cdots q_n z_n$ with $z_i \in B$. At least one of the z_i is not in $\{t_1, t_2\}$ as otherwise t_1, t_2 satisfy a quadratic polynomial over $\mathbb{Q}$. Say $z_i \notin \{t_1, t_2\}$. Let π be a permutation of B which switches z_i with w_i, where $w_i \notin \{t_1, t_2, z_1, \dots, z_n\}$, and fixes all other elements of B (in particular fixes t_1 and t_2). Then $\pi'(t_1 t_2) \neq t_1 t_2$, but $\pi'(t_1) = t_1$, $\pi'(t_2) = t_2$, a contradiction.

Suppose next that $+_{\mathbb{R}}$ were definable over $\mathfrak{R}_m = (\mathbb{R}; \cdot)$. Let $B \subseteq \mathbb{R}^+$ be a maximal set of multiplicitively independent elements, that is, such that $z_1^{q_1} \cdots z_n^{q_n} \neq 1$ for any $z_1, \dots, z_n \in B$ and $q_1, \dots, q_n \in \mathbb{Q}$. Let $\pi\colon B \to B$ be a bijection. First extend π to $\mathbb{R}^+$ by setting $\pi'(z_1^{q_1} \cdots z_n^{q_n}) = \pi(z_1)^{q_1} \cdots \pi(z_n)^{q_n}$, which is easily well-defined and preserves multiplication on $\mathbb{R}^+$. Extend π to all of $\mathbb{R}$ by $\pi(0) = 0$ and $\pi'(-x) = -\pi'(x)$ for $x \in R^+$. π' is still a bijection and a homomorphism from $\mathfrak{R}_m$ to itself. Let $t_1, t_2 \in B$ be algebraically independent elements. Write $t_1 + t_2 = z_1^{q_1} \cdots z_n^{q_n}$ where $z_1, \dots, z_n \in B$. We cannot have that all of the z_i are in $\{t_1, t_2\}$. For if so, then $(t_1 + t_2)^b = t_1^{a_1} t_2^{a_2}$ for some integers a_1, a_2, b with $b \neq 0$. This contradicts the algebraic independence of t_1, t_2. Write $t_1 + t_2 = z_1^{q_1} \cdots z_n^{q_n}$, and we may assume $z_i \notin \{t_1, t_2\}$. Let π switch z_i with $w_i \notin \{t_1, t_2, z_1, \dots, z_n\}$ as before. Then $\pi'(t_1) = t_1$, $\pi'(t_2) = t_2$, but $\pi'(t_1 + t_2) \neq t_1 + t_2$, a contradiction. □

Exercise 9. Let $\mathfrak{R}_z = (\mathbb{R}; +_{\mathbb{R}}, \cdot_{\mathbb{R}}, \mathbb{Z}_{\mathbb{R}})$ be the structure of the reals with the usual operations of addition and multiplication, and a unary relation $\mathbb{Z}_{\mathbb{R}}$ which gives the integers, that is, $\mathbb{Z}_{\mathbb{R}}(a)$ iff $a \in \mathbb{Z}$. Show that $\{\pi\}$ is definable in $\mathfrak{R}_z$.

Remark 1.17. From Tarski's theorem on elimination of quantifiers in the theory of real closed fields, it follows that $\{\pi\}$ is not definable in $\mathfrak{R} = (\mathbb{R}; +, \cdot)$.

1.2. **Prenex normal form.**

We say two wffs (in a language $\mathcal{L}$) of first-order logic, φ and ψ, are *logically equivalent* if $\{\varphi\} \models \psi$ and $\{\psi\} \models \varphi$. Thus, for any structure $\mathfrak{A}$ of $\mathcal{L}$ and any $s\colon \mathrm{Var} \to |\mathfrak{A}|$ we have $\mathfrak{A} \models \varphi[s]$ iff $\mathfrak{A} \models \psi[s]$.

For the purposes of the following discussion, we allow the exists quantifier $\exists$ in the language (officially it is still an abbreviation for $\neg\forall\neg$, but we will now use it as a symbol in the language). Likewise, we use $\wedge$ and $\vee$ as symbols in the language, though they too are still oficially abbreviations.

Definition 1.18. A formula φ is *quantifier free* if it has no occurrence of a quantifier. A quantifier-free formula φ is in *disjunctive normal form* if it is of the form $\varphi = (\varphi_1 \vee \cdots \vee \varphi_n)$, where each φ_i is of the form $\varphi_i = (\psi_{i,1} \wedge \cdots \wedge \psi_{i,m_i})$ and each $\psi_{i,j}$ is either an atomic formula, or the negation of an atomic formula.

Fact 1.19. Every quantifier-free wff φ is logically equivalent to a wff φ' in disjunctive normal form.

Proof. We prove by induction on the wff φ that if φ is quantifier-free then both φ and $\neg\varphi$ are logically equivalent to wffs in disjunctive normal form.

If φ is atomic, then we may take $\varphi' = \varphi$. Also, if the induction hypothesis hold for φ, then it holds for $\neg\phi$ as $\neg\neg\varphi$ is logically equivalent to φ.

It suffices to show that if φ and ψ satisfy the inductive hypothesis, then so does $\varphi \vee \psi$. If φ', ψ' are disjunctive normal forms for φ and ψ, then $\varphi' \vee \psi'$ is already in disjunctive normal form and is logically equivalent to $\varphi \vee \psi$.

It remains to show that $\neg(\varphi \vee \psi)$ can be put into disjunctive normal form. Now $\neg(\varphi \vee \psi)$ is logically equivalent to $(\neg\varphi \wedge \neg\psi)$, and by induction, $\neg\varphi$ and $\neg\psi$ are logically equivalent to formulas in disjunctive normal form. Say $\neg\varphi$ is equivalent to $(\alpha_1 \vee \cdots \vee \alpha_n)$ and $\neg\psi$ is equivalent to $(\beta_1 \vee \cdots \vee \beta_k)$, where the α_i and β_i are conjuncts of atomic formulas and their negations. Then by the usual distributive properties of $\wedge$ and $\vee$ we have

$$(\alpha_1 \vee \cdots \vee \alpha_n) \wedge (\beta_1 \vee \cdots \vee \beta_k) \equiv \bigvee_{i,j} (\alpha_i \wedge \beta_j).$$

where here $\equiv$ means logical equivalence. Each $\alpha_i \wedge \beta_j$ is a conjunction of atomic formulas and their negations, and so $\neg(\varphi \vee \psi)$ is logically equivalent to a fromula in disjunctive normal form.

□

Definition 1.20. A wff φ is in *prenex normal form* if it is of the form $\varphi = Q_1 x_1 \cdots Q_n x_n\ \psi$, where each Q_i is either $\exists$ or $\forall$, and ψ is a quantifier-free formula in disjunctive normal form.

For $\varphi \in \mathrm{Wff}$, let φ^x_z be the result of replacing all the free occurrences of x in φ by z.

We need the following technical fact.

Fact 1.21. For any wff φ, variables x, z with z not occurring in φ, and structure $\mathfrak{A}$ and $s\colon \mathrm{Var} \to |\mathfrak{A}|$ we have that $\mathfrak{A} \models \varphi^x_z[s]$ iff $\mathfrak{A} \models \varphi[s(x|s(z))]$.

Proof. Let $s' = s(x|s(z))$. First, by a straightforward induction on the terms t we show that $s(t^x_z) = s'(t)$. Suppose first that t is a variable. If $t = x$, then the left-hand side is $s(z)$ since $t^x_z = z$. The right-hand side is $(s(x|s(z))(x) = s(z)$. If $t = y \neq x$, then the left-hand side is $s(y)$ and the right-hand side is also $s(y)$ since $(s(x|s(z))(y) = s(y)$. If $t = c$, a constant symbol, then the left-hand side is $s(c) = c^{\mathfrak{A}}$ since $c^x_z = c$, and the right-hand side is also $c^{\mathfrak{A}}$. If $t = f(t_1, \ldots, t_n)$, then the left-hand side is $f^{\mathfrak{A}}(s((t_1)^x_z), \ldots, s((t_n)^x_z))$, which by induction is equal to $f^{\mathfrak{A}}(s'(t_1), \ldots, s'(t_n))$. By definition, this is equal to $s'(t)$.

We now show the fact by induction on φ. First suppose φ is atomic, say $\varphi = R(t_1, \ldots, t_n)$. Then $\mathfrak{A} \models \varphi^x_z[s]$ iff $R^{\mathfrak{A}}(s((t_1)^x_z), \ldots, s((t_n)^x_z))$ since all occurrences of x in φ are free. By the above result for terms, this is equivalent to

$R^{\mathfrak{A}}(s'(t_1), \dots, s'(t_n))$, where again $s' = s(x|s(z))$. By definition, this is equivalent to saying $\mathfrak{A} \models \varphi[s']$.

The boolean case is immediate. So, suppose $\varphi = \forall w\, \psi$, where $w \in \mathrm{Var}$. First suppose that $w \neq x$. Then $\varphi^x_z = \forall w\, \psi^x_z$. So, $\mathfrak{A} \models \varphi^x_z[s]$ iff for all $a \in \mathfrak{A}$ we have $\mathfrak{A} \models \psi^x_z[s(w|a)]$. Note that $w \neq z$ by assumption. By induction, $\mathfrak{A} \models \psi^x_z[s(w|a)]$ iff $\mathfrak{A} \models \psi[(s(w|a))(x|\bar{s}(z))]$, where $\bar{s} = s(w|a)$. This is the same as $\mathfrak{A} \models \psi[(s(w|a))(x|s(z))]$. Since $w \neq x$, this is the same as $\mathfrak{A} \models \psi[(s(x|s(z))(w|a))]$ Saying this holds for all $a \in |\mathfrak{A}|$ is the same as saying $\mathfrak{A} \models \forall w\, \psi[s(x|s(z))]$, that is $\mathfrak{A} \models \varphi[s(x|s(z))]$. Suppose next that $\varphi = \forall x\, \psi$. Then $\varphi^x_z = \varphi$. Also, s and $s(x|s(z))$ agree on all the free variables of φ. So, $\mathfrak{A} \models \varphi^x_z[s]$ iff $\mathfrak{A} \models \varphi[s]$ iff $\mathfrak{A} \models \varphi[s(x|s(z)]$. □

Lemma 1.22. *If $\varphi = Qx\, \psi$ (where $Q = \forall$ or $\exists$), and z is a variable which does not appear in ψ, then φ is logically equivalent to $\varphi' = Qz\, \psi^x_z$, where ψ^x_z is the result of replacing all the free occurrences of x in ψ by z.*

Proof. It is enough to do the case $Q = \forall$. We have $\mathfrak{A} \models \varphi[s]$ iff for every $a \in |\mathfrak{A}|$ we have that $\mathfrak{A} \models \psi[s(x|a)]$. Likewise $\mathfrak{A} \models \varphi'[s]$ iff for every $a \in |\mathfrak{A}|$ we have that $\mathfrak{A} \models \psi^x_z[s(z|a)]$. From Fact 1.21 we have that $\mathfrak{A} \models \psi^x_z[s(z|a)]$ iff $\mathfrak{A} \models \psi[s'(x|s'(z))]$, where $s' = s(z|a)$. Now, z does not appear in ψ, so $\mathfrak{A} \models \psi[s'(x|s'(z))]$ iff $\mathfrak{A} \models \psi[s(x|s'(z)]$. Since $s'(z) = a$, this happens iff $\mathfrak{A} \models \psi[s(x|a)]$. □

Remark 1.23. The formula φ' of Lemma 1.22 is called an *alphabetic variant* of φ.

Theorem 1.24. *Every wff φ in a first-order language is logically equivalent to a wff in prenex normal form.*

Proof. We show by induction on φ that φ and $\neg\varphi$ are logically equivalent to wffs in prenex normal form. For atomic formulas this is immediate. The case $\varphi = \neg\psi$ is also immediate as $\neg\neg\psi$ is equivalent to ψ. The case $\varphi = \forall x\, \psi$ is also immediate. It is enough then to consider the case $\varphi = (\alpha \vee \beta)$.

Say $\alpha = Q_1 x_1 \cdots Q_n x_n\, (\rho)$, and $\beta = R_1 y_1 \cdots R_m y_m\, (\sigma)$, where ρ, σ are in disjunctive normal form. We show by induction on $\max\{n, m\}$ that $(\alpha \vee \beta)$ is equivalent to a wff in prenix normal form. Let $\alpha' = Q_2 x_2 \cdots Q_n x_n\, (\rho)$, and $\beta' = R_2 y_2 \cdots R_m y_m\, (\sigma)$. Then by Lemma 1.22 $Q_1 x_1\, \alpha' \vee R_1 y_1\, \beta'$ is logically equivalent to $Q_1 z_1\, (\alpha')^{x_1}_{z_1} \vee R_1 w_1\, (\beta')^{y_1}_{w_1}$ where z_1, w_1 are distinct variables which do not appear in φ. Now we claim that

$$Q_1 z_1\, (\alpha')^{x_1}_{z_1} \vee R_1 w_1\, (\beta')^{y_1}_{w_1} \equiv Q_1 z_1\, R_1 w_1\, ((\alpha')^{x_1}_{z_1} \vee (\beta')^{y_1}_{w_1})$$

which suffices, as by induction $((\alpha')^{x_1}_{z_1} \vee (\beta')^{y_1}_{w_1})$ is equivalent to a wff in prenix normal form.

To see this, it is enough to observe the following: if z does not occur in a formula ψ, then $(Qz\, \varphi) \vee \psi$ is logically equivalent to $Qz\, (\varphi \vee \psi)$. Fix $\mathfrak{A}$ and $s\colon \mathrm{Var} \to |\mathfrak{A}|$. We consider the case $Q = \forall$, the case $Q = \exists$ being similar.

Suppose first that $\mathfrak{A} \models (\forall z\, \varphi) \vee \psi[s]$. If $\mathfrak{A} \models \psi[s]$ then for all $a \in |\mathfrak{A}|$ we have $\mathfrak{A} \models \psi[s(z|a)]$ as z does not appear in ψ. So, for all $a \in \mathfrak{A}$ we have $\mathfrak{A} \models (\varphi \vee \psi)[s(z|a)]$ and thus $\mathfrak{A} \models \forall z\, (\varphi \vee \psi)$. If $\mathfrak{A} \models \forall z\, \varphi[s]$, then for all $a \in |\mathfrak{A}|$ we have $\mathfrak{A} \models \varphi[s(z|a)]$, and thus $\mathfrak{A} \models (\varphi \vee \psi)[s(z|a)]$. So, $\mathfrak{A} \models \forall z\, (\varphi \vee \psi)[s]$.

Suppose next that $\mathfrak{A} \models \forall z\, (\varphi \vee \psi)[s]$. If $\mathfrak{A} \models \psi[s]$ then $\mathfrak{A} \models (\forall z\, \varphi) \vee \psi[s]$, and we are done. So, assume $\mathfrak{A} \not\models \psi[s]$. Let $a \in |\mathfrak{A}|$. Then $\mathfrak{A} \models (\varphi \vee \psi)[s(z|a)]$. Since $\mathfrak{A} \not\models \psi[s]$, we also have $\mathfrak{A} \not\models \psi[s(z|a)]$ since z does not appear in ψ. So, $\mathfrak{A} \models \varphi[s(z|a)]$. Since $a \in |\mathfrak{A}|$ was arbitrary, we have $\mathfrak{A} \models \forall z\, \varphi[s]$.

Finally, $\neg\varphi \equiv \neg\alpha \wedge \neg\beta \equiv (Q_1'x_1 \cdots Q_n'x_n \neg\rho) \wedge (R_1'y_1 \cdots R_m'y_m \neg\sigma)$, where Q', R' are the dual quantifiers to Q, R. The argument now proceeds just as the above, using the fact that $\neg\sigma$, $\neg\rho$ are equivalent to quantifier-free wffs in disjunctive normal form. □

Example 9. Let $\varphi = \forall x\, \psi(x)$ where $\psi(x) = \exists y\, (x \not\approx y)$. Let $\varphi' = \forall y\, \psi^x_y = \forall y\, \exists y\, (y \not\approx y)$. Then φ is not logically equivalent to φ', so it is important in Lemma 1.22 that the variable z not occur in ψ.

Exercise 10. Let $\varphi = \forall y \exists y\, R(y, y)$. Is φ logically equivalent to $\forall y\, R(y, y)$ or to $\exists y\, R(y, y)$?

Exercise 11. Find a sentence φ in the language of number theory such that $\mathfrak{N} \models \varphi$ and such that every structure that satisfies φ is infinite.

1.3. **Formal Proofs.** We have introduced the notion of logical satisfaction, $\models$ above. This is one of two of the basic notions of implication from first-order logic. The other is the notion of provability, or deducibility. We write $\Gamma \vdash \phi$ for this notion. This notion can be defined in several, ultimately equivalent, ways. The exact definition is not so important for us here, only that this notion satisfies several reasonable properties. However, for the sake of completeness we give one formal definition of $\vdash$. We assume below that the only connectives are $\neg$ and $\rightarrow$.

Definition 1.25. Let φ be a wff in a first-order language $\mathcal{L}$, let $x \in \mathrm{Var}$, and t a term in the language $\mathcal{L}$. We say t is *substitutable* for x in φ if wherever x occurs free in φ, x is not within the scope of a quantifier over a variable that appears in t. More formally, the notion is defined by induction as follows:

(1) If φ is atomic, then t is substitutable for x in φ.
(2) If $\varphi = \neg\alpha$, then t is substitutable for x in φ iff t is substitutable for x in α.
(3) If $\varphi = (\alpha \rightarrow \beta)$, then t is substitutable for x in φ iff t is substitutable for x in α and t is substitutable for x in β.
(4) Suppose $\varphi = \forall z\, \psi$. If $z \neq x$, then t is substitutable for x in φ iff t is substitutable for x in φ and z does not occur in t. If $z = x$, then t is substutable for x in φ.

Definition 1.26. The *logical axioms* (for a given first-order language $\mathcal{L}$) are the universal closures of the following formulas (here α, β, γ are arbitrary formulas of $\mathcal{L}$):

(1) $\alpha \rightarrow (\beta \rightarrow \alpha)$.
(2) $(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma))$
(3) $(\neg\alpha \rightarrow \neg\beta) \rightarrow ((\neg\alpha \rightarrow \beta) \rightarrow \alpha)$
(4) $\forall x\ \alpha \rightarrow \alpha^x_t$ whenever t is *substitutable* for x in ϕ Here α^x_t is the result of replacing x where it occurs free by t.
(5) $\forall x\ (\alpha \rightarrow \beta) \rightarrow (\forall x\ \alpha \rightarrow \forall x\ \beta)$
(6) $\alpha \rightarrow \forall x\ \alpha$, where x is not free in α.
(7) $x_i \approx x_i$
(8) $(x_i \approx x_j) \rightarrow (\alpha \rightarrow \alpha')$, where α is atomic and α' is the result of replacing some of the occurrences of x_i by x_j.

Remark 1.27. The first three categories are just the logical axioms of propositional logic. Note that we always have that $\forall x\, \varphi \vdash \varphi$, as x is always substitutable for x in any formula. Substitutability is necessary in (4) as the next example shows.

Example 10. Let $\varphi = \forall x\, \alpha = \forall x\, (\exists y\, (x \not\approx y))$. Let $t = y$. Then $\alpha^x_t = \exists y\, (y \not\approx y)$. Clearly $\varphi = \forall x\, \alpha$ does not logically imply α^x_t.

Definition 1.28. $\Gamma \vdash \phi$ iff there is a sequence of formulas (a deduction of φ from Γ) $\phi_0, \phi_1, \dots, \phi_n = \phi$ such that for all i either ϕ_i is a logical axiom or an element of Γ, or the formulas ψ, $\psi \to \phi_i$ occur before ϕ_i for some ψ (we say ϕ_i is deduced by modus ponens).

The central result in first-order logic is the Gödel completeness theorem:

Theorem 1.29. *For any first-order language $\mathcal{L}$, set of formulas Γ, and formula ϕ, we have $\Gamma \models \phi$ iff $\Gamma \vdash \phi$.*

As we said, the exact details of the definition of the provability relation are not so important. All we really require is that we be able to prove the completeness theorem and that the notion "$\vec{\phi}$ is a deduction of ψ from Γ" is recursive for recursive Γ (we discuss the notion of recursive below).

We first prove some metatheorems which give information about the $\vdash$ relations, and which correspond to intuitive proof rules in mathematics. The first two, the deduction and contradiction metatheorems are just as in propositional logic.

Lemma 1.30 (Deduction metatheorem)**.** *If $\Gamma \cup \{\varphi\} \vdash \psi$, then $\Gamma \vdash (\varphi \to \psi)$.*

Proof. The proof is just as for propositional logic. We prove the result by induction on the minimal length of a deduction of ψ from $\Gamma \cup \{\varphi\}$. First assume $\psi \in \Gamma \cup \{\varphi\}$. If $\psi \in \Gamma$, then clearly $\Gamma \vdash \psi$ and also $\psi \to (\varphi \to \psi)$ is a logical axiom, so by modus ponens $\Gamma \vdash (\varphi \to \psi)$. If $\psi = \varphi$, then $\Gamma \vdash (\psi \to \psi)$ exactly as in propositional logic.

If ψ is deduced by modus ponens from α and $\alpha \to \psi$, then by induiction $\Gamma \vdash (\varphi \to \alpha)$ and $\Gamma \vdash (\varphi \to (\alpha \to \psi))$. But using group (2) of the logical axioms we have $\varphi \to (\alpha \to \psi) \vdash ((\varphi \to \alpha) \to (\varphi \to \psi))$, and so by modus ponens, $\Gamma \vdash (\varphi \to \psi)$.

Finally, assume ψ is a logical axiom. then $\Gamma \vdash \psi$ and $\varnothing \vdash \psi \to (\varphi \to \psi)$ and so $\Gamma \vdash (\varphi \to \psi)$.

□

Lemma 1.31 (Contradiction metatheorem)**.** *If $\Gamma \cup \{\neg\varphi\} \vdash \alpha, \neg\alpha$, then $\Gamma \vdash \varphi$.*

Proof. Assume $\Gamma \cup \{\neg\varphi\} \vdash \alpha, \neg\alpha$. From the deduction metatheorem we have $\Gamma \vdash (\neg\varphi \to \alpha)$ and $\Gamma \vdash (\neg\varphi \to \neg\alpha)$. From $\Lambda(3)$ we have $\Gamma \vdash \varphi$. □

As with propositional logic, $\Lambda(1) - -(3)$ suffice to show that $\varphi \vdash \neg\neg\varphi$ and $\neg\neg\varphi \vdash \varphi$. Thus we have:

Corollary 1.32. *If $\Gamma \cup \varphi$ is inconsistent, then $\Gamma \vdash \neg\varphi$.*

Lemma 1.33 (Generalization metatheorem)**.** *If $\Gamma \vdash \varphi$ and x is not free in Γ, then $\Gamma \vdash \forall x\, \varphi$.*

Proof. By induction on the minimal length of a deduction of φ from Γ. If $\varphi \in \Gamma$, then x is not free in φ by assumption. By $\Lambda(6)$ we have $\varnothing \vdash (\varphi \to \forall x\, \varphi)$, and it follows that $\Gamma \vdash \forall x\, \varphi$.

If $\varphi \in \Lambda$, then $\forall x\, \varphi$ is also in Λ, and Λ is closed under universal quantification by definition.

If φ is obtained by modus ponens from β, $\beta \to \varphi$, then by induction we have $\Gamma \vdash \forall x\, \beta$ and $\Gamma \vdash \forall x\, (\beta \to \varphi)$. By $\Lambda(5)$ we have $\Gamma \vdash \forall x\, \beta \to \forall x\, \varphi$. By modus ponens, $\Gamma \vdash \forall x\, \varphi$. □

The requirement that x not be free in Γ can by circumvented as the next corollary shows.

Corollary 1.34. *If $\Gamma \vdash \varphi^x_z$ where z is not free in Γ, and z does not occur in φ, then $\Gamma \vdash \forall x\, \varphi$.*

Proof. Since $\Gamma \vdash \varphi^x_z$ and z is not free in Γ, by the generalization metatheorem we have $\Gamma \vdash \forall z\, \varphi^x_z$. But $\forall z\, \varphi^x_z \vdash (\varphi^x_z)^z_x$ since x is always substitutable for z in φ^x_z, by exercise 12. An easy induction shows that $(\varphi^x_z)^z_x = \varphi$. So, $\forall z\, \varphi^x_z \vdash \varphi$. Since x is not free in $\forall z\, \varphi^x_z$, the generalization metatheorem gives that $\forall z\, \varphi^x_z \vdash \forall x\, \varphi$, and so $\Gamma \vdash \forall x\, \varphi$. □

Exercise 12. Show that x is substitutable for z in φ^x_z for any formula φ and variables x, z, with z not in φ. Show also that $(\varphi^x_z)^z_x = \varphi$.

The proof of the completeness theorem will involve introducing some new constant symbols into the language. We need the following metatheorem to be able to eliminate them and get back to deductions in the original language.

Lemma 1.35 (Generalization on constants)**.** *If $\Gamma \vdash \varphi$ and c is a constant symbol not in Γ, then there is a variable y not in φ such that $\Gamma \vdash \forall y\, \varphi^c_y$, and furthermore there is a deduction of $\forall y\, \varphi^c_y$ from Γ in which c does not appear.*

Proof. Let $\alpha_1, \dots, \alpha_n = \varphi$ be a deduction of φ from Γ. Let y be a variable not in any of the α_i. We show by induction on m that $(\alpha_0)^c_y, \dots, (\alpha_m)^c_y$ is a deduction of $(\alpha_m)^c_y$ from Γ.

If $\alpha_m \in \Gamma$, then c does not appear in α_m, so $(\alpha_m)^c_y = \alpha_m$, and so $(\alpha_m)^c_y \in \Gamma$.

If α_m is deduced by modus ponens from α_i and $\alpha_j = \alpha_i \to \alpha_m$, then $(\alpha_j)^c_y = (\alpha_i)^c_y \to (\alpha_m)^c_y$, and $(\alpha_m)^c_y$ is deduced by modus ponens from $(\alpha_i)^c_y$ and $(\alpha_j)^c_y$.

Suppose finally that $\alpha_m \in \Lambda$. We claim that $(\alpha_m)^c_y$ is also in Λ.

Claim 3. If $\alpha \in \Lambda$ and y is a variable not in α, then $\alpha^c_y \in \Lambda$.

Proof. For α in Λ groups (1), (2), (3), (5), or (7) this is clear.

Suppose $\alpha \in \Lambda(4)$, say $\alpha = \forall x\, \beta \to \beta^x_t$ where t is substitutable for x in β. Then $\alpha^c_y = \forall x\, \beta^c_y \to (\beta^x_t)^c_y$. Now, for any $\beta \in$ Wff, t substitutable for x in β, and y not in β, $(\beta^x_t)^c_y = (\beta^c_y)^x_{t^c_y}$ (this can be formally proved by a straightforward induction on β). So, $\alpha^c_y = \forall x\, \beta^c_y \to (\beta^c_y)^x_{t^c_y}$. Finally, x occurs free in β^c_y exactly where x occurs free in β. Also, t^c_y is substitutable for x in β^c_y as t^c_y has the same variables as t except perhaps the extra variable y, but y does not appear in β, so does not appear as a quantified variable in β^c_y. Thus, $\alpha^c_y \in \Lambda(4)$.

Suppose next that $\alpha \in \Lambda(6)$, say $\alpha = (\beta \to \forall x\, \beta)$, where x is not free in β. Then $\alpha^c_y = (\beta^c_y \to \forall x\, \beta^c_y)$. Since x is not free in β, we also have that x is not free in β^c_y (since $y \neq x$ as y does not occur in α). Thus $\alpha^c_y \in \Lambda(6)$.

Finally, suppose $\alpha \in \Lambda(8)$. Say $\alpha = (x_i \approx x_j \to (\beta \to \beta'))$, where β' is obtained from β (an atomic formula) by replacing some of the occurrences of x_i by x_j. Then α^c_y is the formula $x_i \approx x_j \to (\beta^c_y \to (\beta')^c_y)$. Since $y \neq x_i, x_j$, $(\beta')^c_y$ is obtained from β^c_y by replacing the same occurrence of x_i by x_j as for β. So, $\alpha^c_y \in \Lambda(8)$. □

This also completes the proof of Lemma 1.35.

□

Finally, before starting the proof of the completeness theorem, we show that suitable alphabetic variants which are provably equivalent to the original formula always exist.

Lemma 1.36. *Let $\varphi \in$ Wff, $x \in$ Var, and t be a term. There is a formula φ' with the same free variables as φ such that t is substitutable for x in φ', and $\varphi \vdash \varphi'$, $\varphi' \vdash \varphi$.*

Proof. By induction on φ. If φ is atomic, we can let $\varphi' = \varphi$. The boolean cases are easy. Suppose $\varphi = \forall w \, \psi$ for some variable w. If $w = x$ then we may take $\varphi' = \varphi$, as x is not free in φ in this case. So, assume $w \neq x$. Consider $\bar{\varphi} = \forall z \, \psi^w_z$ where z is a variable not appearing in φ or t. We claim that $\varphi \vdash\dashv \bar{\varphi}$. To see that $\varphi \vdash \bar{\varphi}$, by the generalization metatheorem it is enough to see that $\varphi \vdash \psi^w_z$ as z is not free in φ. Since z is substitutable for w in ψ (as does not appear in φ) from $\Lambda(4)$ we have that $\varphi \vdash \psi^w_z$. To see that $\bar{\varphi} \vdash \varphi$, by the generalization metatheorem it is enough to show that $\bar{\varphi} \vdash \psi$, since w is not free in $\bar{\varphi}$. Since z does not appear in ψ, all the occurrrences of z in ψ^w_z are free and furthermore w is substitutable for z in ψ^w_z. So by $\Lambda(4)$, $\bar{\varphi} \vdash (\psi^w_z)^z_m = \psi$.

Finally, by induction let $\varphi' = \forall z \, (\psi^w_z)'$, where $(\psi^w_z)' \vdash\dashv \psi^w_x$ and t is substitutable for x in $(\psi^w_z)'$. Since t is substitutable for x in $(\psi^w_z)'$ and z does not occur in t, t is substitutable for x in $\forall z \, (\psi^w_z)' = \varphi'$. From the generalization metatheorem we immediately have that $\forall z \, \psi^w_z \vdash\dashv \forall z \, (\psi^w_z)'$.

□

1.4. **Proof of the completeness theorem.** Let $\mathcal{L}$ be a language of first-order logic. Let $\Gamma \subseteq \mathrm{Wff}_{\mathcal{L}}$ and suppose $\Gamma \models \varphi$. We show $\Gamma \vdash \varphi$. If $\Gamma \nvdash \varphi$, then by the contradiction metatheorem, $\Gamma \cup \{\neg\varphi\}$ is consistent. If we can show that every consistent set of wffs is satisfied by some $\mathfrak{A}$ and $s \colon \mathrm{Var} \to |\mathfrak{A}|$, then there is a $\mathfrak{A}$ and s which satisfies $\Gamma \cup \{\neg\varphi\}$. This is impossible, however, as $\mathfrak{A} \models \Gamma$, and thus $\mathfrak{A} \models \varphi[s]$.

So, it suffices to show the following.

Theorem 1.37. *If Γ is a consistent set of wffs in a first-order language $\mathcal{L}$, then there is a structure $\mathfrak{A}$ for $\mathcal{L}$ and a $s \colon Var \to |\mathfrak{A}|$ which satisfies Γ.*

Proof. Let $\mathcal{L}'$ be the language $\mathcal{L}$ together with countably many new constant symbols c_i which do not occur in $\mathcal{L}$.

Let $\varphi_0, \varphi_1, \dots$ enumerate all the existential formulas of $\mathcal{L}'$, that is, all formulas of the form $\varphi_i = \exists y_i \; \psi_i$ for some variable y_i and formula ψ_i. For each φ_i choose one of the new constant symbols c_{φ_i} not in φ_i and not in φ_j nor equal to c_{φ_j} for any $j < i$. So, for each existential formula $\varphi = \exists y \; \psi$ we have the constant symbol c_φ defined. Let Γ' be Γ together with all formulas of the form $\exists y \, \psi \to \psi(c_\varphi)$ for each existential formula $\varphi = \exists y \, \psi$. Here $\psi(c_\varphi)$ abbreviates $\psi^y_{c_\varphi}$. The new formulas $\sigma_i = \exists y_i \, \psi_i \to \psi_i(c_{\varphi_i})$ are called the *Henkin formulas*, and the c_φ the *Henkin witnesses*. So,

$$\Gamma' = \Gamma \cup \{\exists y_i \psi_i \to \psi_i(c_{\varphi_i})\}_{i \in \omega}.$$

Claim 4. Γ' is consistent.

Proof. If not, let n be least such that $\Gamma_n = \Gamma \cup \{\sigma_0, \dots, \sigma_n\}$ is inconsistent. Note that Γ is still consistent with respect to the language $\mathcal{L}'$. This follows from Lemma 1.35. Namely, if $\Gamma \vdash_{\mathcal{L}'} (\alpha \wedge \neg\alpha)$, where $\alpha \in \mathrm{Wff}_{\mathcal{L}'}$, then by Lemma 1.35 we may replace the new constant symbols in α by variables to get an $\alpha' \in \mathrm{Wff}_{\mathcal{L}}$ and such that $\Gamma \vdash (\alpha' \wedge \neg\alpha')$ by a proof in $\mathcal{L}$, a contradiction.

So, Γ_{n-1} is consistent ($\Gamma_{-1} = \Gamma$), and $\Gamma_{n-1} \cup \{\sigma_n\}$ is inconsistent. Let $\sigma_n = \exists y\, \psi \to \psi(c_\varphi)$, where $\varphi_n = \exists y\, \psi$. So, c_φ does not appear in φ nor in $\sigma_0, \dots, \sigma_{n-1}$. By the contradiction metatheorem, $\Gamma_{n-1} \vdash \neg\sigma_n$. Now, for any wffs α and β we easily have $\neg\alpha \vdash (\alpha \to \beta)$ (using the deduction metatheorem), and so $\neg(\alpha \to \beta) \vdash \alpha$ by the contradiction metatheorem. Thus, $\Gamma_{n-1} \vdash \exists y\, \psi$. Also, $\neg(\alpha \to \beta) \vdash \neg\beta$ by the contradiction metatheorem as $\beta \vdash (\alpha \to \beta)$ by a logical axiom. So, $\Gamma_{n-1} \vdash \neg\psi(c_\varphi)$. Since c_φ does not appear in Γ_{n-1}, from Lemma 1.35 we have that $\Gamma_{n-1} \vdash \forall z\, \neg(\psi^y_{c_\varphi})^{c_\varphi}_z$ for some variable z (which we may assume which does not occur in φ), and furthermore c_φ does not occur in this deduction. Since c_φ does not occur in ψ, $(\psi^y_{c_\varphi})^{c_\varphi}_z = \psi^y_z$. So, $\Gamma_{n-1} \vdash \forall z\, \neg\psi^y_z$. Since z did not occur in ψ, z occurs free in ψ^y_z exactly where y appeared free in ψ, and y is substitutable for z in ψ^y_z. So, by $\Lambda(4)$ we have $\forall z\, \neg\psi^y_z \to \neg(\psi^y_z)^z_y$. But, $(\psi^y_z)^z_y = \psi$, so $\forall z\, \neg\psi^y_z \vdash \neg\psi$. Since y is not free in $\forall z\, \neg\psi^y_z$, by the generaliztion metatheorem we have $\forall z\, \neg\psi^y_z \vdash \forall y\, \neg\psi$. Thus, $\Gamma_{n-1} \vdash \forall y\, \neg\psi$. Since $\Gamma \vdash \exists y\, \psi = \neg\forall y\, \neg\psi$, we see that Γ_{n-1} is inconsistent, a contradiction. □

We next enlarge Γ' to Γ'' which a maximal consistent set of wffs in the language $\mathcal{L}'$. This is done exactly as for propositional logic. If the language $\mathcal{L}$ is countable (and so $\mathcal{L}'$ is also countable) we enumerate the wffs of $\mathcal{L}'$ as $\alpha_0, \alpha_1, \dots$, set $\Gamma_{-1} = \Gamma$, and then inductively define $\Gamma_n = \Gamma_{n-1} \cup \{\alpha_n\}$ if this set is consistent, and otherwise $\Gamma_n = \Gamma_{n-1}$. Then $\Gamma'' = \bigcup_m \Gamma_n$ is a maximal consistent set. If the language $\mathcal{L}$ (and so $\mathcal{L}'$) is uncountable, then we use AC to wellorder the wffs and use the same argument, taking unions at limit stages.

As in the proof for propositional logic, the maximality of Γ'', along with the contradiction metatheorem, gives that for every wff α in $\mathcal{L}'$, either $\alpha \in \Gamma''$ or $\neg\alpha \in \Gamma''$ (and both cannot hold). Also, if $\Gamma'' \vdash \alpha$, then $\alpha \in \Gamma''$.

Finally, from Γ'' we build a structure $\mathfrak{A}$ and map $s\colon \mathrm{Var} \to |\mathfrak{A}|$. Let T be the set of terms in the language $\mathcal{L}'$. We define a relation $\equiv$ on T by $t \equiv u$ iff $(t \approx u) \in \Gamma''$.

Claim 5. $\equiv$ is an equivalence relation on T.

Proof. We have $\varnothing \vdash \forall x\, (x \approx x)$ by $\Lambda(7)$, and so by $\Lambda(4)$ we have $\varnothing \vdash t \approx t$ for any term t. So $t \equiv t$ that is, $\equiv$ is reflexive.

We have $\varnothing \vdash \forall x\, \forall y\, (x \approx y \to y \approx x)$ by an application of $\Lambda(8)$ [Consider the atomic formula $x \approx x$, so $(x \approx y) \to ((x \approx x) \to (y \approx x)) \in \Lambda$. Since $\varnothing \vdash x \approx x$, it easily follows that $\varnothing \vdash (x \approx y) \to (y \approx x)$.] From $\Lambda(4)$ it follows that $\varnothing \vdash (t \approx u \to u \approx t)$. So, if $t \equiv u$, that is $(t \approx u) \in \Gamma''$, then $\Gamma'' \vdash u \approx t$, and so $(u \approx t) \in \Gamma''$, and so $u \equiv t$. So, $t \equiv u$ implies $u \equiv t$, so $\equiv$ is symmetric.

Suppose $t \equiv u$ and $u \equiv v$, so $(t \approx u)$ and $(u \approx v)$ are in Γ''. It suffices to show that $\varnothing \vdash \forall x\, \forall y\, \forall z\, (x \approx y) \to ((y \approx z) \to (x \approx z))$ since by $\Lambda(4)$ (choosing x, y, z to be variables not in t, u, v) it follows that $\varnothing \vdash (t \approx u) \to ((u \approx v) \to (t \approx v))$, and so $\Gamma'' \vdash (t \approx v)$. This however follows easily from $\Lambda(8)$ by considering the atomic formula $y \approx z$, and replacing the occurrence of y by x as in $\Lambda(8)$, which gives $\varnothing \vdash (y \approx x) \to ((y \approx z) \to (x \approx z))$. Since $\varnothing \vdash (x \approx y \to y \approx x)$ by the above, we have that $\varnothing \vdash (x \approx y) \to ((y \approx z) \to (x \approx z))$. □

We let $[t]$ denote the equivalence class of t under the relation $\equiv$, for t a term in $\mathcal{L}'$.

We let $|\mathfrak{A}| = \{[t] \colon t \in T\}$. We let $s(x) = [x]$ for $x \in \mathrm{Var}_{\mathcal{L}'} = \mathrm{Var}_{\mathcal{L}}$. for c a constant symbol of $\mathcal{L}'$, we let $c^{\mathfrak{A}} = [c]$. It remains to define the interpretations $R^{\mathfrak{A}}$, $f^{\mathfrak{A}}$ of the relation and function symbols.

We define $R^{\mathfrak{A}}([t_1], \dots, [t_n])$ iff $R(t_1, \dots, t_n) \in \Gamma''$, and similarly for functions symbols we define $f^{\mathfrak{A}}([t_1], \dots, [t_n]) = [f(t_1, \dots, t_n)]$, however, we need to show these definitions are welldefined. Suppose $t_1 \equiv u_1, \dots, t_n \equiv u_n$ and $R^{\mathfrak{A}}([t_1], \dots, [t_n])$, that is, $R(t_1, \dots, t_n) \in \Gamma''$. From the generalization metatheorem, the deduction metatheorem, and $\Lambda(8)$, we have that

$$\begin{aligned}\varnothing \vdash \forall y_1\, \forall z_1 \cdots \forall y_n\, \forall z_n\, ((y_1 \approx z_1) \to \cdots (y_n \approx z_n) \to\\ (R(y_1, \dots, y_n) \to R(z_1, \dots, z_n)) \cdots).\end{aligned}$$

We may choose the variables y_i, z_i so that they do not appear in the t_i or u_i. From $\Lambda(4)$ we get that

$$\varnothing \vdash ((t_1 \approx u_1) \to \cdots (t_n \approx u_n) \to (R(t_1, \dots, t_n) \to R(u_1, \dots, u_n)) \cdots).$$

Since $\Gamma'' \vdash t_i \approx u_i$, we have that $\Gamma'' \vdash R(u_1, \dots, u_n)$. So, $R(u_1, \dots, u_n) \in \Gamma''$. The argument for functions is similar.

At this point we have a well-defined structure $\mathfrak{A} = (|\mathfrak{A}|; R_i^{\mathfrak{A}}, f_i^{\mathfrak{A}}, c_i^{\mathfrak{A}})$ and $s \colon \mathrm{Var}_{\mathcal{L}'} \to |\mathfrak{A}|$.

Claim 6. For every wff φ of $\mathcal{L}'$, we have $\mathfrak{A} \models \varphi[s]$ iff $\varphi \in \Gamma''$.

Proof. By induction on φ. If φ is atomic of the form $\varphi = R(t_1, \dots, t_n)$, then $\mathfrak{A} \models \varphi[s]$ iff $R^{\mathfrak{A}}([t_1], \dots, [t_n])$ which by definition of $R^{\mathfrak{A}}$ holds iff $R(t_1, \dots, t_n) \in \Gamma''$. If $\varphi = (t \approx u)$, then $\mathfrak{A} \models \varphi[s]$ iff $s(t) = s(u)$. By definition of s, this means $[t] = [u]$, that is, $t \equiv u$. By definition of $\equiv$, this happens iff $(t \approx u) \in \Gamma''$.

The boolean case is just as in propositional logic. For example, say $\varphi = (\alpha \to \beta)$. Then $\mathfrak{A} \models \varphi[s]$ iff $\mathfrak{A} \not\models \alpha[s]$ or $\mathfrak{A} \models \beta[s]$. By induction this is equivalent to saying $\alpha \notin \Gamma''$ or $\beta \in \Gamma''$. By maximality of Γ'' this holds iff $\neg\alpha \in \Gamma''$ or $\beta \in \Gamma''$. As Γ'' is closed under deduction, this is equivalent to $(\alpha \to \beta) \in \Gamma''$.

Suppose that $\varphi = \forall x\, \psi$. First assume that $\varphi \in \Gamma''$, and we show that $\mathfrak{A} \models \varphi[s]$. Let $[t] \in |\mathfrak{A}|$, we must show that $\mathfrak{A} \models \psi[s(x|[t])]$. Let ψ', by Lemma 1.36, be an alphabetic variant of ψ, that is, ψ' is provably equivalent to ψ and t is substitutable for x in ψ'. Then $\mathfrak{A} \models \psi[s(x|[t])]$ iff $\mathfrak{A} \models \psi'(s(x|[t])$ iff $\mathfrak{A} \models (\psi')^x_t[s]$. The last equivalence is by the following fact, which is an extension of Fact 1.21.

Fact 1.38. Let $\varphi \in \mathrm{Wff}$, $x \in \mathrm{Var}$, and t a term which is substitutable for x in φ. Then $\mathfrak{A} \models \varphi^x_t[s]$ iff $\mathfrak{A} \models \varphi[s(x|s(t)]$.

Proof. The proof is essentially identical to that of Fact 1.21. □

By induction, $\mathfrak{A} \models (\psi')^x_t[s]$ iff $(\psi')^x_t \in \Gamma''$. As $\varphi = \forall x\, \psi \in \Gamma''$, we have $\forall x\, \psi' \in \Gamma''$, as $\varphi = \forall x\, \psi \vdash \forall x\, \psi'$ (from the generalization metatheorem, using the fact that $\psi \vdash \psi'$). Since t is substitutable for x in ψ', by $\Lambda(4)$ we have that $\forall x\, \psi' \vdash (\psi')^x_t$, and so $(\psi')^x_t \in \Gamma''$. Thus, $\mathfrak{A} \models \forall x\, \psi$.

Suppose $\mathfrak{A} \models \forall x\, \psi[s]$. We must show that $\forall x\, \psi \in \Gamma''$. If not, then by maximality of Γ'' we have that $\neg\forall x\, \psi \in \Gamma''$, so $\exists x\, \neg\psi \in \Gamma''$. But $\exists x\, \neg\psi \to \neg\psi^x_c$ is one of the Henkin formulas of Γ'. So, $\Gamma'' \vdash \neg\psi^x_c$. Here c is one of the new constant symbols of $\mathcal{L}'$ which in particular does not appear in ψ. Since $\mathfrak{A} \models \forall x\, \psi[s]$, $\mathfrak{A} \models \psi[s(x|[c])]$. since c is substitutable for x in ψ, we have $\mathfrak{A} \models \psi^x_c[s]$. By induction, $\psi^x_c \in \Gamma''$, a contradiction.

□

This completes the proof of Theorem 1.37.

□

1.5. **Applications of the completeness theorem.** As with propositional logic, the completeness theorem immediately implies the compactness theorem, which we state next. The compactness theorem can also be proved directly (without mentioning the completeness theorem or $\vdash$) using ultraproducts.

Theorem 1.39 (Compactness theorem)**.** *Let $\mathcal{L}$ be a language of first-order logic, and $\Gamma \subseteq Wff_{\mathcal{L}}$. Then Γ is satisfiable (i.e., there is an $\mathcal{L}$-structure $\mathfrak{A}$ and a $s\colon Var \to |\mathfrak{A}|$ with $\mathfrak{A} \models \varphi[s]$ for all $\varphi \in \Gamma$) iff every finite $\Gamma_0 \subseteq \Gamma$ is satisfiable.*

Proof. Supose every finite $\Gamma_0 \subseteq \Gamma$ is satisfiable. If Γ is not satisfiable, then $\Gamma \models \alpha$ for any wff α, so $\Gamma \models \alpha \wedge \neg\alpha$ for some (any) wff α. By the completeness theorrem, $\Gamma \vdash \alpha \wedge \neg\alpha$ for some α, that is, Γ is inconsistent. Since proofs are finite, we have $\Gamma_0 \vdash \alpha \wedge \neg\alpha$ for some finite $\Gamma_0 \subseteq \Gamma$. But then $\Gamma_0 \models \alpha \wedge \neg\alpha$, a contradiction as Γ_0 is satisfiable. □

Exercise 13. Let $\mathcal{L}$ be a first-order language, and $\Gamma \subseteq \mathrm{Wff}_{\mathcal{L}}$. Show that if Γ has finite models of arbitrarily large size, then Γ has an infinite model. [hint: add to Γ a collection of wffs $\{\psi_n\}_{n\in\omega}$ which together force $|\mathfrak{A}|$ to be infinite.]

Exercise 14. Let $\mathcal{L} = \{\cdot\}$ be the language of group theory. Let Γ_0 be the (finite) set of axioms for a group. Show that there does not exists a set $\Gamma \supseteq \Gamma_0$ such that a group G satisfies Γ iff G is finite. In particular, there is no single formula (in the language of group theory) which can "say" that G is a finite group. [hint: use the previous exercise.]

We saw before that there is a set of wffs Γ (in the empty language) such that a structure $\mathfrak{A}$ satisfies Γ iff $|\mathfrak{A}|$ is infinite. From Exercise 13 if follows that there is no set of wffs Γ in the empty language such that $\mathfrak{A} \models \Gamma$ iff $|\mathfrak{A}|$ is finite.

We next use the compactness theorem to construct non-isomorphic models of the theory of the natural numbers. This will produce a non-standard model of arithmetic.

Definition 1.40. Let $\mathcal{L}$ be a first-order languaqge, and $\mathfrak{A}$ an $\mathcal{L}$-structure. We let $\mathrm{Th}(\mathfrak{A})$, the *theory of* $\mathfrak{A}$, denote the set of sentences φ in $\mathcal{L}$ such that $\mathfrak{A} \models \varphi$.

Theorem 1.41. *Let $\mathcal{L}$ be the language of number theory. Let $Th(\mathfrak{N})$ be the theory of $\mathfrak{N}$ for the language $\mathcal{L}$. Then there is a countable model of $Th(\mathfrak{N})$ which is not isomorphic to $\mathfrak{N}$.*

Proof. Let $\mathcal{L}' = \mathcal{L} \cup \{c\}$ where c is a new constant symbol in the languagge. Let $\Gamma' = \Gamma \cup \{\varphi_n\}_{n\in\omega}$ where $\varphi_n = (c > S^{n-1}(0))$. Any finite subset Γ'_0 of Γ' is satisfiable, by letting $s(c)$ to some element of $\mathbb{N}$ larger than all the numbers k such that $s^{k-1}(0)$ is mentioned in Γ'_0. By the compactness theorem, Γ' is satisfiable, say by $\mathfrak{A}'$. Then $c^{\mathfrak{A}'} > s^k(0)$ for all k, and thus $\mathfrak{A}'$ is not isomorphic to $\mathfrak{N}$ as there is no such element in $\mathfrak{N}$ (note that any isomorphism must fix all the $S^k(0)$). We may take $\mathfrak{A}'$ to be countable by taking an elementary substructure of $\mathfrak{A}$ which is countable and contains $c^{\mathfrak{A}'}$. Alternatively, we may note that the proof of the completeness theorem produces a countable $\mathfrak{A}'$, as $\mathcal{L}'$ is countable. □

Remark 1.42. We should regard Theorem 1.45 more formally as a theorem of some background metatheory in which the standard $\mathfrak{N}$ is defined, and in which the various notions, such as countable models, can be defined. For example, we could take the metatheory to be ZFC, which defines $\mathbb{N}$ as the set of finite ordinals, and permits the proof of Theorem 1.45 to be formalized.

Remark 1.43. Theorem 1.45 also holds, by the same proof, if we just assume that the language $\mathcal{L}$ contains the language $\mathcal{L}_N$ of number theory, and we interpret the conclusion as saying that $\mathfrak{N}$ is not isomorphic to $\mathfrak{N}'_{\mathcal{L}_N}$, where $\mathfrak{N}'_{\mathcal{L}_N}$ denotes the restriction of $\mathfrak{N}'$ to the language $\mathcal{L}_N$ of number theory.

Definition 1.44. We say two structures $\mathfrak{A}$, $\mathfrak{B}$ for a first-order langiage $\mathcal{L}$ are *elementary equivalent*, written $\mathfrak{A} \equiv \mathfrak{B}$, if for any sentence φ of $\mathcal{L}$ we have that $\mathfrak{A} \models \varphi$ iff $\mathfrak{B} \models \varphi$.

As an immediate consequence of Theorem 1.45, it follows that we cannot "axiomatize" the natural numbers in the sense of having a set of formulas Γ in the language of number theory (or any larger language) such that $\mathfrak{N}$ is the unique countable structure up to isomorphism which satisfies Γ.

We refer to the structures $\mathfrak{N}'$ for $\mathrm{Th}(\mathfrak{N})$ which are not isomorphic to $\mathfrak{N}$ as *non-standard* models of arithmetic. We refer to $\mathfrak{N}$ as the "standard model" (see Remark 1.42). It follows immediately from the definition of a non-standard model that if $\mathfrak{N}'$ is a non-standard model of arithmetic, then $\mathfrak{N}' \equiv \mathfrak{N}$, where $\mathfrak{N}$ is the standard model.

The next theorem describes the basic structure of a non-standard model of arithmetic. Let $\mathfrak{N}'$ be a non-standard model of arithmetic. By a "$\mathbb{Z}$-chain" in $\mathfrak{N}'$ we mean a subset $\{a_i\}_{i\in\mathbb{Z}}$ of $|\mathfrak{N}'|$ such that for all $i \in \mathbb{Z}$, there is no element $x \in |\mathfrak{N}'|$ with $a_i < x < a_{i+1}$ (here $<=<^{\mathfrak{N}'}$ denotes the ordering of $\mathfrak{N}'$).

Theorem 1.45. *Let $\mathfrak{N}'$ be a non-standard model of arithmetic. Then $\mathfrak{N}'$ has an initial segment isomorphic to the standard model $\mathfrak{N}$. The elements of $\mathfrak{N}'$ not in this intial segment (the "infinite" elements of $\mathfrak{N}'$) form a dense set of $\mathbb{Z}$-chains (that is, between any two $\mathbb{Z}$-chains there is a third $\mathbb{Z}$-chain) with no smallest or largest $\mathbb{Z}$-chain.*

Proof. We use repeatedly the fact that $\mathfrak{N}' \equiv \mathfrak{N}$. First, $0^{\mathfrak{N}}$ is the least element of $\mathfrak{N}$, that is, $\mathfrak{N} \models \forall x\,(0 < x)$. So, $\mathfrak{N}'$ satisfies the same sentence, so $0^{\mathfrak{N}'}$ is the least element of $\mathfrak{N}'$. Similarly $S(0)^{\mathfrak{N}}$ is the least element of $\mathfrak{N}$ greater than $0^{\mathfrak{N}}$. That is, $\mathfrak{N} \models \forall x\,(0 < x \rightarrow (x \approx S(0) \vee S(0) < x))$. Since $\mathfrak{N}'$ satisfies the same formula, we see that $S(0)^{\mathfrak{N}'}$ is the least element of $\mathfrak{N}'$ greater than $0^{\mathfrak{N}'}$ in $\mathfrak{N}'$. Continuing in this manner we see that the map $\pi : \mathbb{N} \rightarrow \mathbb{N}'$ given by $\pi(S^k(0)^{\mathfrak{N}}) = S^k(0)^{\mathfrak{N}'}$ is an order-preserving bijection between $\mathbb{N} = |\mathfrak{N}|$ and an initial segment of $\mathbb{N}' = |\mathfrak{N}'|$. Now, π must also be a homomorphism from $\mathfrak{N}$ to $\mathfrak{N}'$. For example, $\mathfrak{N} \models (S^k(0) + S^\ell(0) \approx S^{k+\ell}(0))$, so $\mathfrak{N}'$ satisifes this sentence. So $(S^k(0))^{\mathfrak{N}'} + (S^\ell(0))^{\mathfrak{N}'} = (S^{k+\ell}(0))^{\mathfrak{N}'}$, that is, $\pi(S^k(0)) + \pi(S^\ell(0)) = \pi(S^{k+\ell}(0))$.

Let $I(\mathbb{N}')$ be the elements of $\mathbb{N}'$ not in the range of π. These are the "infinite" elements of $\mathbb{N}'$. Since π is an isomorphism between $\mathfrak{N}$ and $\mathrm{ran}(\pi)$, we may identitfy $\mathrm{ran}(\pi)$ with $\mathfrak{N}$, and just write $\mathbb{N}' = \mathbb{N} \cup I$, where $\mathbb{N}$ is an initial segment of $\mathbb{N}'$. That is, $k < c$ for all $k \in \mathbb{N}$ and $c \in I$.

It remains to show that I consists of a dense set of $\mathbb{Z}$-chains. First note that if $c, d \in I$ then $c + d \in I$, and if $c, d \neq 0$ then $c \cdot d \in I$. $\mathfrak{N} \models \forall x\,\forall y\,(x > 0 \wedge y > 0 \rightarrow (x + y > x \wedge x \cdot y \geqslant x))$, so $\mathfrak{N}'$ satisfies the same sentence. Since $c, d > 0$, $c + d > c > S^k(0)$ for any k, so $c + d \in I$. Likewise $c \cdot d \in I$. Since

$\mathfrak{N} \models (\forall x\,(x > 0 \to \exists y\,(y + S(0) \approx x \wedge \forall w\,\neg(y < w < x))$, $\mathfrak{N}'$ satisfies the same sentence, so c has an immediate predecessor in $\mathfrak{N}'$. We denote this predecessor by $c - S(0)$. Also, $\mathfrak{N} \models \forall x\,((x + S(0) > S^{k+1}(0)) \to x > S^k(0))$, and thus, since $c > S^{k+1}(0)$ for all k, we have that $c - S(0) > S^k(0)$ for all k, so $c - S(0) \in I$. Continuing, we see that the $\mathbb{Z}$-chain $\{c \pm S^k(0)\}_{k \in \mathbb{Z}} \subseteq I$.

We show that there is a $\mathbb{Z}$-chain between c and d. Asssume $c < d$ and c, d are in distinct $\mathbb{Z}$-chains. $\mathfrak{N} \models \forall x\,\exists y\,(y + y \approx x \vee y + y \approx x + S(0))$, so this sentence also holds in $\mathfrak{N}'$. So, let $y \in \mathbb{N}'$ be such that $y + y = c + d$ or $y + y = c + d + 1$. For example, say $y + y = c + d + 1$. The equation $y > c + S^k(0)$ is equivalent to $y + y > c + c + S^{2k}(0)$. But, $y + y = c + d + 1 > c + c + S^{2k}(0)$ as $d > c + S^{2k}(0)$. We use here the fact that statements hold in $\mathfrak{N}$ iff they hold in $\mathfrak{N}'$.

A similar argument shows that there is a y such that $y + y = c$ or $y + y = c + 1$, and the $\mathbb{Z}$-chain of y is less than the $\mathbb{Z}$-chain of c. Likewise, the $\mathbb{Z}$-chain of $d + d$ is greater than the $\mathbb{Z}$-chain of d. □

Remark 1.46. Any two countable dense linear orderings without endpoints are order-isomnorphic (isomorphic ti $\mathbb{Q}$), so the order-type of I, for $\mathfrak{N}'$ a countable non-standard model, is uniquely determined, as $\mathbb{Q}$ copies of $\mathbb{Z}$.