Contents

1 Languages, Structures, and Completeness **2**
1.1 Languages 2
1.2 Structures 6
1.3 Completeness 9

2 Embeddings and Lindström's Classification of First-Order Logic **13**
2.1 Embeddings 13
2.1.1 Theorems of Löwenheim and Skolem 17
2.1.2 Games of Ehrenfeucht and Fraïssé[10][12] 19
2.2 Lindström's Classification of First-Order Logic 22
2.2.1 Comparing logics[18] 23
2.2.2 Lindström's theorem[18] 24

3 Boole, Arrow, and Compactness Revisited **26**
3.1 Boole 26
3.2 Arrow's Theorem[3][15][16][19] 30
3.3 Ramsey's results[4][10] 32
3.4 The Ultraproduct Construction 35
3.4.1 Reduced Products[10] 35
3.4.2 Ultraproducts 37
3.5 Non-standard Universes 40
3.5.1 Hyperreal Numbers[13] 40
3.5.2 Non-standard Calculus 42

Chapter 1

Languages, Structures, and Completeness

1.1 Languages

In natural language, we have a collection of symbols and rules to manipulate the symbols in order to communicate. The process of mathematical formalization begins with the notion of a *logic* $\mathfrak{L}$, which dictates the nature of the information that can be presented. The prime object of our discussion will be the *first-order logic* $\mathfrak{L}_{\omega\omega}$, which essentially puts some limits on what we can write.[1] A *formal language* $\mathcal{L}$ consists of logical constants and non logical symbols- constants and variables, that can be plugged into the 'templates' provided in the logic to convey information. A language establishes the smallest collection of 'labels' needed to sufficiently describe a mathematical structure. Like words in natural languages, labels don't have any interpretation until we associate them with one.

Logical constants of interest to us will be elements in the set $\{=, \wedge, \vee, \neg, \forall, \exists\}$ where the symbols carry their usual meaning. These are symbols whose semantic value does not change with interpretation of the language.[11] For convenience, we define some special combinations of the logical constants which occur frequently:

$$(a \to b) := (\neg a) \vee b$$
$$(a \leftrightarrow b) := (a \to b) \wedge (b \to a)$$

[1] See Section 2.2 for a detailed discussion.

Remark. Observe that the logical constants we mentioned are related to each other as

$$(a \vee b) \Leftrightarrow \neg(\neg a \wedge \neg b),$$
$$\exists a\ b \Leftrightarrow \neg(\forall a\ \neg b).$$

Hence, for the purposes of this text we will only prove results about one of the two logical connectives $\{\vee, \wedge\}$ and one of the two existential quantifiers $\{\exists, \forall\}$ and the result will be true for the others as long as closure under negation holds. However, the distinction becomes necessary later, for instance when dealing with *"statements with universal quantification"*.

Non logical variables are symbols $v_0, v_1, v_2, \ldots$ which are manipulated without referring to their actual value. We sometimes use $x, y, z, \ldots$ for variables.

Definition 1.1.[12] The *signature* or the *vocabulary* is the set σ of non logical constant symbols of the language along with their descriptions. It is a union of

(i) σ_R, the set of *relation- symbols*, along with a function $\sigma_R \to \mathbb{N}$ which associates each $R \in \sigma_R$ to some natural number called its *arity*.

(ii) σ_F, the set of *function-symbols*, along with a function $\sigma_F \to \mathbb{N}$ which associates each $f \in \sigma_F$ to its *arity*.

(iii) σ_C, the set of *constant-symbols*.

Remark. Based on different conventions, equality '=' can be classified as both a logical constant, and a binary relation symbol. For conciseness of definitions coming up, we shall assume $= \in \sigma_R$ for all languages σ and has arity 2, and shall not state it explicitly when writing the signatures.

Every time we mention a language $\mathcal{L}$ it is implied that we are dealing with both, a signature and an underlying logic. We write $\mathfrak{L}(\sigma)$ for the language formed from the logic $\mathfrak{L}$ and the signature σ.

Example 1.2. If we want to talk about groups, we will need one function-symbol $\cdot$ of arity 2 (binary) to label the group operation and one constant-symbol e to label the group identity. Thus $\sigma_g = \{\cdot, e\}$ suffices as a signature for groups. Write $\mathcal{L}_g$ for the first-order language of groups $\mathfrak{L}_{\omega\omega}(\sigma_g)$.

Example 1.3. $\sigma_{\exp} = \{<, +, \times, \exp, 0, 1\}$ is a signature that lets us describe the ordered set of real numbers under exponentiation, where $<$ is a binary relation-symbol, $+$ and $\times$ are binary function-symbols and 0 and 1 are constant-symbols. Write $\mathcal{L}_{\exp}$ for the first-order language $\mathfrak{L}_{\omega\omega}(\sigma_{\exp})$

Just like words can be formed from certain combinations of the letters of the alphabet, a *term* is a sequence of symbols which can actually have a meaningful interpretation. For example, $+(0, \exp(1))$ and $\exp(+(\times(y,\ 0),\ x))$ are $\mathcal{L}_{\exp}$-terms (written more conveniently as $0 + \exp(1)$ and $\exp((y \times 0) + x)$ respectively.)

Definition 1.4.[12] For a language $\mathcal{L}$, *the set of $\mathcal{L}$-terms* is defined as the smallest set $\mathrm{Term}(\mathcal{L})$ such that

(i) variables $v_0, v_1, v_2, ... \in \mathrm{Term}(\mathcal{L})$

(ii) The set of constant $\mathcal{L}$-symbols $\mathcal{L}_C \subset \mathrm{Term}(\mathcal{L})$

(iii) if f is an n-ary function $\mathcal{L}$-symbol and $t_1, t_2, \dots, t_n \in \mathrm{Term}(\mathcal{L})$ then $f(t_1, t_2, \dots, t_n) \in \mathrm{Term}(L)$

In a natural language, words allow us to make assertions. In this context, assertions will involve stating how terms are related. This is achieved by means of *formulae.* ϕ is an *atomic formula* if ϕ is $R(t_1, t_2, \dots, t_n)$ where the relation-symbol R is n-ary. Call the set of atomic $\mathcal{L}$-formulae $\mathrm{Atom}(\mathcal{L})$. These form the 'building blocks' of all formulae. For example, $\phi : (0 < exp(x + y))$ and $\psi : (1 = 0)$ are elements of $\mathrm{Atom}(\mathcal{L}_{\exp})$.

Definition 1.5.[12] The set of *well-formed formulae* of a language $\mathcal{L}$ is the smallest set $\mathrm{Form}(\mathcal{L})$ such that

(i) The set of atomic $\mathcal{L}$-formulae $\mathrm{Atom}(\mathcal{L}) \subset \mathrm{Form}(\mathcal{L})$

(ii) $\neg\phi \in \mathrm{Form}(\mathcal{L})$ whenever $\phi \in \mathrm{Form}(\mathcal{L})$

(iii) $\phi \wedge \psi \in \mathrm{Form}(\mathcal{L})$ whenever ϕ and $\psi \in \mathrm{Form}(\mathcal{L})$

(iv) $\forall x\ \phi \in \mathrm{Form}(\mathcal{L})$ whenever $\phi \in \mathrm{Form}(\mathcal{L})$

i.e. the set of well formed formulae is the smallest set containing all atomic formulae that is closed under negation, conjunction and quantification. Observe that the set automatically becomes closed under $\vee$ and $\exists$. For example, $\exists x\ (y < x) \wedge (y < 1) \in \mathrm{Form}(\mathcal{L}_{\exp})$. Note that the formulae are simply assertions with no truth value associated- for in fact 1 and $<$ are simply labels without any interpretation as of now.

Definition 1.4 and Definition 1.5 are *inductive*: any set containing all $\mathcal{L}$-constants and variables and satisfying property (iii) must coincide with $\mathrm{Term}(\mathcal{L})$, and likewise for $\mathrm{Form}(\mathcal{L})$. This is useful when proving results about terms and formulae. We look at another inductive definition which will be useful:

Definition 1.6.[12] For a languate $\mathcal{L}$, the *quantifier depth*[2] of an $\mathcal{L}$-formula ϕ is depth(ϕ) such that

(i) $\text{depth}(\phi) = 0$ if ϕ is an atomic $\mathcal{L}$-formula

(ii) $\text{depth}(\neg\phi) = \text{depth}(\phi)$

(iii) $\text{depth} \bigwedge \Phi = \max\{\text{depth}(\phi) : \phi \in \Phi\}$

(iv) $\text{depth}(\forall x\, \phi(x)) = \text{depth}(\phi) + 1$

We look at a useful result:

Lemma 1.7.[12] *For a finite vocabulary, for each n and l, there are only finitely many formulas in l variables of depth at most n, up to equivalence.*

Proof. We first show this for depth 0:
Since the vocabulary is finite with finitely many variables, we can only have finitely many atomic formulae $\{\psi_0, \psi_1, \psi_2, ..., \psi_k\}$. If ϕ is a boolean combination of these, then there is K a collection of subsets of $\{0, 1, ..., k\}$ such that

$$\phi \leftrightarrow \bigvee_{X \in K} \Big(\bigwedge_{i \in X} \psi_i \wedge \bigwedge_{i \notin X} \neg\psi_i \Big)$$

so we can have at most 2^{2^k} depth 0 formulae.
Formulae of depth $n+1$ are simply boolean combinations of formulae of form $\forall x\, \phi$ where $\text{depth}(\phi) \leq n$ so result follows by induction. □

A *free variable* is a variable not bound by any quantifier. A variable that isn't free is a *bound variable*. If $x, y, \ldots, z$ are the free variables occuring in a formula ϕ, then we write $\phi(x, y, \ldots, z)$; likewise for terms. It can be said that ϕ makes assertions *about* whatever is assigned to the variable-symbols $x, y, \ldots., z$.[3]

A *sentence* is a formula with no free variables. Sentences are assertions about the entire universe. Examples of $\mathcal{L}_{\text{exp}}$-sentences are $\exists x\, \neg(x = 0)$ (intended to mean *'there is some non-zero element'* upon proper interpretation) and $\forall x\, (1 < x) \wedge (x < 0)$ (intended to mean *'every element is greater than 1 and smaller than 0'* upon proper interpretation.) A *theory* is a set of sentences.

[2] also called *quantifier rank*
[3] See [12] for a more detailed discussion.

Example 1.8.

$$\begin{aligned}\{&\forall x\,\exists y\ \exp(y) = x,\\ &\forall x\,\forall y\ \exp(x) < \exp(y) \leftrightarrow x < y,\\ &\exp(0) = 1\ \}\end{aligned}$$

is an $\mathcal{L}_{\exp}$-theory.

We have thus established a framework to express mathematical ideas which compares well with intuitive ideas behind natural languages. However, the very first definition- that of logical constants is one that has been subject matter for philosophical debate over many years, for, in the words of one of the founding fathers of model theory Alfred Tarski himself,

> No objective grounds are known to me which permit us to draw a sharp boundary between [logical and non-logical expressions].[6]

We shall however brush these issues aside since they will not affect our study significantly.

1.2 Structures

Now that we have established a language, we would want to assign meaning to the symbols- the function-symbols should label some actual functions, the relation-symbols should label some actual relations. We do this by means of an operation $\theta \mapsto \theta^{\mathcal{M}}$ which tells us how to interpret a symbol in the signature σ as some object in a mathematical *structure* $\mathcal{M}$.

Definition 1.9.[12] A *structure* $\mathcal{M}$ is a *realization* of the language $\mathcal{L} = \mathfrak{L}(\sigma)$ if it has

(i) a set M, called the *universe* of $\mathcal{M}$,

(ii) a function $f^{\mathcal{M}} : M^n \to M$ for each n-ary function- symbol $f \in \sigma$,

(iii) a relation $R^{\mathcal{M}} \subset M^n$ for each n-ary relation-symbol $R \in \sigma$,

(iv) an element $c^{\mathcal{M}} \in M$ for each constant-symbol $c \in \sigma$.

We write the structure as a tuple[4] $\mathcal{M} = (M, f^{\mathcal{M}}, \ldots, R^{\mathcal{M}}, \ldots, c^{\mathcal{M}}, \ldots)$ and call it an $\mathcal{L}$-*structure*.

[4]Not necessarily finite

Example 1.10. $\mathcal{N} = (\mathbb{N}, +, 5)$ is an $\mathcal{L}_g$-structure, where $\cdot^{\mathcal{N}} = +$ and $e^{\mathcal{N}} = 5$ Note that this is not a group, however all that matters is that every symbol in the language has an assigned meaning- the $\mathcal{L}_g$-term $t(x) = \cdot(x, e)$ is interpreted as $t^{\mathcal{N}}(x) = x + 5$, and the $\mathcal{L}_g$- sentence ϕ given by $\cdot(e, e) = e$ corresponds to $\phi^{\mathcal{N}}$: $5 + 5 = 5$.

This also gives us a nice interpretation of terms: they can be viewed as functions taking in tuples from the universe corresponding to free variables and producing a single element as output. So in the above example, $t^{\mathcal{N}}$ is a function $\mathbb{N} \to \mathbb{N}$ given by $x \mapsto x+5$. Formulae, on the other hand, take in tuples from the universe and output *true* or *false* based on whether the formula holds for the tuple. For instance the $\mathcal{L}_g$ formula $\phi(x):\ (x \cdot e) \cdot e = (x \cdot x) \cdot e$ corresponds to $\phi^{\mathcal{N}}(x)$ which asks if $(x+5)+5 = (x+x)+5$. Clearly, $\phi^{\mathcal{N}}(5)$ is true while $\phi^{\mathcal{N}}(2)$ is not.

We need some notion of truth of formulae in structures. We say an $\mathcal{L}$-structure $\mathcal{M}$ *satisfies* or *models* an $\mathcal{L}$-formula $\phi(\bar{x})$ for an n-tuple $\bar{a} \in M^n$ if $\phi^{\mathcal{M}}(\bar{a})$ is true in $\mathcal{M}$, and write $\mathcal{M} \vDash \phi(\bar{a})$. The formal definition is inductive on Form($\mathcal{L}$):

Definition 1.11.[12] For a signature $\mathcal{L}$, an $\mathcal{L}$-structure $\mathcal{M} = (M, ...)$ and an $\mathcal{L}$-formula $\phi(\bar{x})$,

(i) If $\phi(\bar{x})$ is an atomic $\mathcal{L}$-formula, then ϕ is of the form $R(\bar{x})$ for some n-ary relation-symbol $R \in \mathcal{L}$. For $\bar{a} \in M^n$ write $\mathcal{M} \vDash \phi(\bar{a})$ iff $\bar{a} \in R^{\mathcal{M}}$.

(ii) If ϕ is $\neg\psi$ then $\mathcal{M} \vDash \phi$ iff $\mathcal{M} \nvDash \psi$.

(iii) If ϕ is $\psi \wedge \theta$ then $\mathcal{M} \vDash \phi$ iff $\mathcal{M} \vDash \psi$ and $\mathcal{M} \vDash \theta$.

(iv) If $\phi(\bar{x})$ is $\forall y\ \psi(\bar{x}, y)$ then $\mathcal{M} \vDash \phi(\bar{a})$ iff $\mathcal{M} \vDash \psi(\bar{a}, b)$ whenever $b \in M$.

The cases for $\vee$ and $\exists$ follow from these base cases so are not stated separately.

The notion of a structure satisfying a theory follows immediately: an $\mathcal{L}$-structure $\mathcal{M}$ *satisfies* an $\mathcal{L}$-theory T if $\mathcal{M} \vDash \phi$ whenever $\phi \in T$. We say $\mathcal{M}$ is a *model* of T, written $\mathcal{M} \vDash T$. A theory is said to be *satisfiable* if it has atleast one model.

A class of $\mathcal{L}$-structures $\mathcal{K}$ is an *elementary class* if there is an $\mathcal{L}$-theory T such that $\mathcal{K} = \{\mathcal{M} : \mathcal{M} \vDash T\}$.
T is called the set of *axioms* of $\mathcal{K}$ and we say T *axiomatizes* $\mathcal{K}$.

Example 1.12. Consider the $\mathcal{L}_g$-theory T_g given by

$$\begin{aligned} T_g = \{ &\forall x,\ e \cdot x = x \\ &\forall x\, \exists y,\ x \cdot y = e \\ &\forall x\, \forall y\, \forall z,\ x \cdot (y \cdot z) = (x \cdot y) \cdot z\ \}. \end{aligned}$$

This theory is satisfiable- $\mathcal{Z} = (\mathbb{Z}, +, 0)$ models it when $\cdot^{\mathcal{Z}} = +$, $e^{\mathcal{Z}} = 0$. In fact, any group is a model of this theory- the class of groups is an elementary class axiomatized by T_g. If we look at $T_{a.g.} = T_g \cup \{\forall x\, \forall y,\ x \cdot y = y \cdot x\}$, this precisely axiomatizes the elementary class of abelian groups. One can check that $\mathcal{Z} \vDash T_{a.g.}$.

Example 1.13. Define $\sigma_r = \{+, \cdot, 0, 1\}$, and $\mathcal{L}_r = \mathfrak{L}_{\omega\omega}(\sigma)$ the language of *rings*. Then the $\mathcal{L}_r$ theory

$$\begin{aligned} T_r = \{\ &\forall x,\ x + 0 = x, \\ &\forall x\, \exists y,\ x + y = 0, \\ &\forall x\, \forall y\, \forall z,\ (x + y) + z = x + (y + z), \\ &\forall x,\ x \cdot 1 = x, \\ &\forall x\, \forall y\, \forall z,\ (x \cdot y) \cdot z = x \cdot (y \cdot z), \\ &\forall x\, \forall y\, \forall z,\ (x + y) \cdot z = x \cdot z + y \cdot z, \\ &\forall x\, \forall y\, \forall z,\ x \cdot (y + z) = x \cdot y + x \cdot z\ \} \end{aligned}$$

axiomatizes the elementary class of rings. $T_{c.r.} = T_r \cup \{\forall x\, \forall y,\ x \cdot y = y \cdot x\}$ gives the axioms of *commutative rings*, while $T_f = T_{c.r.} \cup \{\forall x\, \exists y,\ x = 0 \vee x \cdot y = 1\}$ forms the set of *field* axioms. *Can we axiomatize finite fields using first order logic?*[5]

Observe we have used the same language $\mathcal{L}_r$ of rings to write theories for structures that are much richer. We could, for instance, encode all of these using a different signature $\sigma_f = \sigma_r \cup \{-\}$ (and $\mathcal{L}_f = \mathfrak{L}_{\omega\omega}(\sigma_f)$ where $-$ is a unary operation-symbol to be interpreted as additive inverse. This theory would allow us to axiomatize fields using one less sentence (existence of additive inverse would be trivially true.) However, from a strictly model theoretic perspective, fields as $\mathcal{L}_f$-structures are different from fields as $\mathcal{L}_r$-structures. *Can we define fields as $\mathcal{L}_{f*}$-structures where $\mathcal{L}_{f*}$ is formed by adding to the signature of $\mathcal{L}_f$ a unary operation-symbol $^{-1}$ to be interpreted as multiplicative inverse?*

[5] See Corollary 1.17 and then try answering this question.

Definition 1.14.[9] For a language $\mathcal{L}$, the *full theory* of an $\mathcal{L}$-structure $\mathcal{M}$ is the set $\mathrm{Th}(\mathcal{M})$ of all $\mathcal{L}$-sentences satisfied by $\mathcal{M}$.

For a language $\mathcal{L} = \mathfrak{L}(\sigma)$, given a model $\mathcal{M}$ and $\bar{m}$ the string of all the elements in M, we create a new signature $\sigma_{\mathcal{M}} = \sigma \cup \{c : c^{\mathcal{M}} \in M\}$, i.e. by adding constant-symbols for every element in M. Then for the language $\mathcal{L}_{\mathcal{M}} = \mathfrak{L}(\sigma_{\mathcal{M}})$, $(\mathcal{M}, \bar{m})$ is an $\mathcal{L}_{\mathcal{M}}$-structure by interpreting the constant-symbols with $\bar{m}$. We do this so that if $\phi(\bar{x})$ is an $\mathcal{L}$-formula and $\bar{a} \in \mathcal{M}$, then $\phi(\bar{a})$ can now be seen as an $\mathcal{L}_{\mathcal{M}}$-sentence. This allows us more freedom in the sets we can define.

The *atomic diagram* of $\mathcal{M}$ is the set of all quantifier-free $\mathcal{L}_{\mathcal{M}}$-sentences satisfied by $(\mathcal{M}, \bar{m})$. Quantifier-free $\mathcal{L}_{\mathcal{M}}$-sentences correspond to atomic $\mathcal{L}$-formulae, or negations thereof.

$$\mathrm{Diag}(\mathcal{M}) = \{\phi(\bar{a}) : \ \phi(\bar{x}) \text{ or } \neg\phi(\bar{x}) \in \mathrm{Atom}(\mathcal{L}),\ \bar{a} \in M^n,\ \mathcal{M} \vDash \phi(\bar{a})\}$$

The *elementary diagram* of $\mathcal{M}$ is the set of all $\mathcal{L}_{\mathcal{M}}$-sentences satisfied by $(\mathcal{M}, \bar{m})$.

$$\mathrm{Diag}_{\mathrm{el}}(\mathcal{M}) = \{\phi(\bar{a}) : \ \phi(\bar{x}) \in \mathrm{Form}(\mathcal{L}),\ \bar{a} \in M^n,\ \mathcal{M} \vDash \phi(\bar{a})\} \ = \mathrm{Th}(\mathcal{M}, \bar{m})$$

Needless to say, $\mathcal{M} \vDash \mathrm{Th}(\mathcal{M})$, $\mathcal{M} \vDash \mathrm{Diag}(\mathcal{M})$, and $\mathcal{M} \vDash \mathrm{Diag}_{\mathrm{el}}(\mathcal{M})$.

1.3 Completeness

In the first two sections, we discussed separately the notions of *syntax* and *semantics.* One of the greatest achievement of mathematical logic is the unification of the two.

A *proof* of an $\mathcal{L}$-formula ϕ from an $\mathcal{L}$-theory T is a finite sequence of $\mathcal{L}$-formulae $(\psi_0, \psi_1, ..., \psi_n)$ such that $\psi_n = \phi$ and for each $i \in \{0, 1, ..., n\}$, either $\psi_i \in T$ or ψ_i follows from $\{\psi_0, \psi_1, ..., \psi_{i-1}\}$ using rules of simple inference (for example q follows from $\{p \to q, p\}$, and u follows from $\{u \wedge v\}$). We write $T \vdash \phi$ if there is a proof of ϕ from T.[12]

An $\mathcal{L}$-theory T is *complete* if for any $\mathcal{L}$-sentence ϕ, exactly one of $T \vdash \phi$ or $T \vdash \neg\phi$ is true. Observe that the full theory of any $\mathcal{L}$-structure $\mathcal{M}$ is complete, hence is also called *the complete theory of* $\mathcal{M}$.

An $\mathcal{L}$-theory T is *inconsistent* if there exists an $\mathcal{L}$-sentence ϕ such that

$T \vdash (\phi \wedge \neg\phi)$. Otherwise, the theory is *consistent.* A consistent theory does not prove any contradiction. This idea becomes very powerful when combined with the idea that our proof system is sound (see Theorem 1.15), and is what our proof system relies on.

All of the above are completely syntactic notions. We move on to discuss the semantic notion of *logical consequence*: An $\mathcal{L}$-sentence ϕ is a logical consequence of an $\mathcal{L}$-theory T if for any $\mathcal{L}$-structure $\mathcal{M}$, $\mathcal{M} \models \phi$ whenever $\mathcal{M} \models T$. We write $T \models \phi$.

Theorem 1.15.[9] ***(Soundness)*** *For a first-order*[6] *language $\mathcal{L}$, for any $\mathcal{L}$-theory T and $\mathcal{L}$-sentence ϕ, $T \models \phi$ whenever $T \vdash \phi$.*

Proof. (Sketch) We proceed by induction on the length n of the proof of ϕ. If $n = 1$ then $\phi \in T$ so $T \models \phi$. If we assume the theorem holds till n, then for ϕ having proof of length $n+1$ say $(\psi_0, \psi_1, ..., \psi_n)$, we must have ψ_n following from $\{\psi_0, \psi_1, ..., \psi_{n-1}\}$ using simple rules of inference. For example if ψ_n follows from $\{\psi_j = (\psi_i \rightarrow \psi_n),\ \psi_i\}$ then ψ_i and ψ_j have proofs of lengths $i+1 < n$ and $j+1 < n$ respectively, so by induction hypothesis $T \models \psi_i$ and $T \models \psi_j$. But from definition of $\models$, $T \models \psi_i \rightarrow \psi_n$ iff $T \models \psi_n$ or $T \not\models \psi_i$. Hence we must have $T \models \psi_n$, i.e. $T \models \phi$.
We similarly consider other rules of inference to complete the proof. □

In other words, every satisfiable theory is consistent. This tells us that our proof system is *sound*, i.e. we cannot prove something that is not a logical consequence. However, can we prove everything that *is* a logical consequence? This was shown in one of the most remarkable results about logic, proven by Kurt Gödel in 1929 who reduced the problem to deal only with special syntactic forms using an ad-hoc argument. The result was later proven by Leon Henkin in 1949 by direct construction of a model (see [10] for details of the proof.) We state the theorem:

Theorem 1.16.[10] ***(Gödel's Completeness Theorem)*** *For a first-order language $\mathcal{L}$, for any $\mathcal{L}$-theory T and $\mathcal{L}$-sentence ϕ, $T \vdash \phi$ whenever $T \models \phi$.*

This says every consistent theory is satisfiable. Coupled with soundness, it formally establishes the equivalence of the semantic notion of logical consequence and the finitist syntactic notion of proof.

A keen observer might have noticed that we *proved* Theorem 1.15 to show

[6] See Section 2.2.

that it is semantically true in our model of mathematics. Does this mean that before writing the proof, we had no idea that the proof would actually mean the theorem is true? Since the theorem itself connects syntactic proofs and logical consequences, have we in some sense done both- proving the statement and also showing that the proof is actually meaningful- in the same text? No. A language cannot prove results about itself. To avoid running into such self-referential situations, it is important that we make a clear distinction between the *object language*: the language about which we are proving results, and the *meta-language*: the language which we are using to prove the result. The two don't need to be related to each other at all, and we generally assume that the meta-language is well behaved, so that we don't have to worry about its consistency.

An $\mathcal{L}$-theory T is *finitely satisfiable* if every finite subset of T is satisfiable. It is trivially true that every satisfiable theory is finitely satisfiable, since $T \vDash \Delta$ for every $\Delta \subseteq T$. The converse, however, is non-trivial and forms the cornerstone of model theory:

Corollary 1.17.[12] ***(Compactness Theorem)*** *For a first-order language $\mathcal{L}$, an $\mathcal{L}$-theory is satisfiable if and only if it is finitely satisfiable.*

Proof. Any satisfiable theory is clearly finitely satisfiable. To show the converse, let $\mathcal{L}$-theory T be finitely satisfiable. If T is not satisfiable, then by Theorem 1.16, T is inconsistent i.e. $T \vdash (\phi \wedge \neg\phi)$ for some $\mathcal{L}$-sentence ϕ. Let σ be a proof of $(\phi \wedge \neg\phi)$ from T. Since proofs are finite, σ uses sentences from some finite subset $\Delta \subseteq T$, hence $\Delta \vdash (\phi \wedge \neg\phi)$. But then Δ is an inconsistent, hence unsatisfiable finite subset of T, a contradiction. $\square$

Perhaps at this point you are worried the proof of Corollary 1.17 relies on Theorem 1.16, the proof of which we have not stated. Since this result is so important, we shall present a more self-contained proof in Section 3.4.2.

We look at some interesting consequences of the theorem:

Example 1.18. Any $\mathcal{L}$-theory T with arbitrarily large finite models has an infinite model. To show this, consider the sentence

$$\psi_n : \exists x_0\, \exists x_1, ... \exists x_{n-1},\ \neg(x_0 = x_1) \wedge \neg(x_0 = x_2) \wedge ... \wedge \neg(x_{n-2} = x_{n-1})$$

for $n \in \mathbb{N}$. This sentence asserts *there are atleast n distinct elements.* For any finite subset $\{n_0, n_1, ..., n_k\} \subset \mathbb{N}$, we have a model of $T \cup \{\psi_{n_0}, \psi_{n_1}, ..., \psi_{n_k}\}$ by choosing an appropriately large model of T. Then $T \cup \{\psi_n : n \in \mathbb{N}\}$ is finitely satisfiable (since T is finitely satisfiable), and thus by compactness theorem there is an $\mathcal{L}$-structure $\mathcal{M} \vDash T \cup \{\psi_n : n \in \mathbb{N}\} \Rightarrow \mathcal{M} \vDash T$, but $\mathcal{M}$ must be infinite. For instance, this tells us there cannot be any first-order theory of finite sets (or groups) i.e. a theory that is satisfied iff the set (or group) is finite.

Example 1.19. The signature required to describe graphs $\sigma_{\mathrm{gr}} = \{R\}$ has a single binary relation-symbol, to be interpreted as a non-reflexive symmetric relation. Suppose the first-order language of graphs $\mathcal{L}_{\mathrm{gr}} = \mathfrak{L}_{\omega\omega}(\sigma_{\mathrm{gr}})$ allowed us to write a theory T which would axiomatize all cyclic graphs, i.e. $\mathcal{M} \vDash T$ if and only if $\mathcal{M}$ is a cyclic graph. Then for the $\mathcal{L}_{\mathrm{gr}}$-sentence

$$\phi_n : \quad \exists x_0 \, \exists x_1, ..., \exists x_n, \bigwedge_{i=0}^{n-2} x_i \, R \, x_{i+1},$$

the theory $T^* = T \cup \{\phi_n : n \in \mathbb{N}\}$ is finitely satisfied by choosing an appropriately large cyclic graph. Hence by Corollary 1.17, it is satisfiable, but any model of T^* must contain an infinite chain (and hence be acyclic). Thus an acyclic graph satisfies T, a contradiction and cyclicity is not a first-order definable property.

Chapter 2

Embeddings and Lindström's Classification of First-Order Logic

2.1 Embeddings

How are various structures related to each other? There can be various ways to classify a structure as being the 'same' as, or 'contained' within some other structure. The notion of a 'homomorphism' occurs frequently in mathematics- maps which preserve some kind of structure. We generalize the notion of a homomorphism to structures.

For a language $\mathcal{L}$, a map $\theta : \mathcal{M} \to \mathcal{N}$ is an $\mathcal{L}$-*homomorphism* if it is a map $M \to N$ that preserves the interpretation of all symbols in $\mathcal{L}$:[9]

(i) For all n-ary $f \in \mathcal{L}_F$ and $\bar{a} \in M^n$, $\theta(f^{\mathcal{M}}(\bar{a})) = f^{\mathcal{N}}(\theta(\bar{a}))$,

(ii) For all n-ary $R \in \mathcal{L}_R$ and $\bar{a} \in M^n$, $\bar{a} \in R^{\mathcal{M}}$ iff $\theta(\bar{a}) \in R^{\mathcal{N}}$,

(iii) For all $c \in \mathcal{L}_C$, $\theta(c^{\mathcal{M}}) = c^{\mathcal{N}}$.

An injective $\mathcal{L}$-homomorphism is an $\mathcal{L}$-*embedding.* If there is an $\mathcal{L}$-embedding $\mathcal{M} \to \mathcal{N}$, we say $\mathcal{M}$ is a *substructure* of $\mathcal{N}$, and $\mathcal{N}$ is an *extension* of $\mathcal{M}$, written $\mathcal{M} \subseteq \mathcal{N}$.

Note that the $\mathcal{L}_g$-substructure of a group is not necessarily a subgroup[1]-

[1] However for $\mathcal{L}_g^* = \mathcal{L}_g \cup \{^{-1}\}$, the $\mathcal{L}_g^*$-substructures of groups are indeed subgroups if we interpret the unary function-symbol $^{-1}$ with taking the inverse.

for example, $(\mathbb{N}_0, +, 0) \subseteq (\mathbb{Z}, +, 0)$. This is because embeddings say nothing about satisfaction of theories. We need stronger notions of containment for that.

We say an embedding $\theta : \mathcal{M} \to \mathcal{N}$ is an *elementary embedding* if $\mathcal{M} \vDash \phi(\bar{a}) \Leftrightarrow \mathcal{N} \vDash \phi(\theta(\bar{a}))$ for all $\phi \in \mathrm{Form}(\mathcal{M})$, $\bar{a} \in M$. If such an elementary embedding exists from $\mathcal{M}$ to $\mathcal{N}$ exists then we say $\mathcal{M}$ is an *elementary substructure* of $\mathcal{N}$, and $\mathcal{N}$ is an *elementary extension* of $\mathcal{M}$; written $\mathcal{M} \preccurlyeq \mathcal{N}$. While extensions just add more elements to the structure, elementary extensions do this while preserving all first order properties. The following proposition makes this clear:

Theorem 2.1.[12] *For a language $\mathcal{L}$ and $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{N}$, the following are true:*

(i) $\mathcal{N} \vDash \mathrm{Diag}(\mathcal{M}) \Leftrightarrow \mathcal{M} \subseteq \mathcal{N}$

(ii) $\mathcal{N} \vDash \mathrm{Diag}_{\mathrm{el}}(\mathcal{M}) \Leftrightarrow \mathcal{M} \preccurlyeq \mathcal{N}$

Proof. Consider $\mathcal{N}$ as an $\mathcal{L}_{\mathcal{M}}$-structure.

(i) If $\mathcal{N} \vDash \mathrm{Diag}(\mathcal{M})$ then consider $\theta : \mathcal{M} \to \mathcal{N}$ given by $\theta(m) = m^{\mathcal{N}}$. θ is an injection: If $m_1 \neq m_2$ then $\neg(m_1 = m_2)$ is in $\mathrm{Diag}(\mathcal{M})$ so $\mathcal{N} \vDash \neg(m_1 = m_2)$, and so $\theta(m_1) \neq \theta(m_2)$.

If $f \in \mathcal{L}_F$ and $\bar{m} \in M$ such that $f(\bar{m}) = n$ then $(f(\bar{m}) = n)$ is in $\mathrm{Diag}(\mathcal{M})$ so $f(\theta(m)) = \theta(n)$. Likewise for relation-symbols and constant-symbols
Thus θ is an $\mathcal{L}$-embedding so $\mathcal{M} \subseteq \mathcal{N}$.

Conversely, if $\mathcal{M} \subseteq \mathcal{N}$ with the embedding given by θ and $\phi \in \mathrm{Atom}(\mathcal{M})$, then for $\bar{m} \in m$ we have

For variable-symbol x, $\theta(x^{\mathcal{M}}) = x^{\mathcal{N}}$.
For constant-symbol c, $\theta(c^{\mathcal{M}}) = c^{\mathcal{N}}$.
For $\bar{t}$ a sequence in $\mathrm{Term}(\mathcal{L})$ and $f \in \mathcal{L}_F$, we have

$$\theta(f^{\mathcal{M}}(\bar{t}^{\,\mathcal{M}})) = f^{\mathcal{N}}(\theta(\bar{t}^{\,\mathcal{M}})) = f^{\mathcal{N}}(\bar{t}^{\,\mathcal{N}})$$

since θ is an $\mathcal{L}$-embedding. Thus we have shown by induction that for $t \in \mathrm{Term}(\mathcal{L})$, $\theta(t^{\mathcal{M}}) = t^{\mathcal{N}}$.

If $\phi \in \mathrm{Atom}(\mathcal{L})$ then $\phi(\bar{x})$ is $R(\bar{t})$ for $R \in \mathcal{L}_R$ and $\bar{t}(\bar{x}) \in \mathrm{Term}(\mathcal{L})$. Then

$$\begin{aligned}\mathcal{M} \vDash \phi(\bar{m}) &\Leftrightarrow \bar{t}^{\,\mathcal{M}}(\bar{m}^{\,\mathcal{M}}) \in R^{\mathcal{M}} \\ &\Leftrightarrow \theta(\bar{t}^{\,\mathcal{M}}(\bar{m}^{\,\Updownarrow})) \in R^{\mathcal{N}} \\ &\Leftrightarrow \bar{t}^{\,\mathcal{N}}(\bar{m}^{\mathcal{N}}) \in R^{\mathcal{N}} \\ &\Leftrightarrow \mathcal{N} \vDash \phi(\bar{m})\end{aligned}$$

and so $\mathcal{N} \vDash \mathrm{Diag}(\mathcal{M})$.

(ii) If $\mathcal{N} \vDash \mathrm{Diag}_{\mathrm{el}}(\mathcal{M})$ then θ above is an elementary embedding. Conversely, if $\mathcal{M} \preccurlyeq \mathcal{N}$ with the elementary embedding given by θ then from above, $(N) \vDash \mathrm{Diag}(\mathcal{M})$, but θ being elementary can be used to show closure under $\forall, \neg$ and $\wedge$ so $\mathcal{N} \vDash \mathrm{Diag}_{\mathrm{el}}(\mathcal{M})$.

$\square$

Part (i) says *quantifier free formulae are preserved under superstructure.* (Preservation under substructure is trivially true.)
We now look at an important criterion for determining elementary containment:

Theorem 2.2.[12] ***(Tarski-Vaught Test)*** *For a language $\mathcal{L}$, and $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{N}$ such that $\mathcal{M} \subseteq \mathcal{N}$, we have $\mathcal{M} \preccurlyeq \mathcal{N}$ if and only if for any $\phi(x) \in \mathrm{Diag}_{\mathrm{el}}(\mathcal{M})$, we have $\mathcal{N} \vDash \exists x\, \phi(x) \Rightarrow \exists m \in M,\ \mathcal{N} \vDash \phi(m)$.*

Proof. If $\mathcal{M} \preccurlyeq \mathcal{N}$ then $\mathcal{N} \vDash \exists x\, \phi(x) \Rightarrow \mathcal{M} \vDash \exists x\, \phi(x) \Rightarrow \exists m \in M,\ \mathcal{M} \vDash \phi(m)$ and hence $\mathcal{N} \vDash \phi(m)$. Conversely, if for any $\phi(x) \in \mathrm{Diag}_{\mathrm{el}}(\mathcal{M})$, we have $\mathcal{N} \vDash \exists x\, \phi(x) \Rightarrow \exists m \in M,\ \mathcal{N} \vDash \phi(m)$, then

if $\phi(\bar{a})$ is quantifier free then $\mathcal{M} \vDash \phi(\bar{a}) \Leftrightarrow \mathcal{N} \vDash \phi(\bar{a})$

$\mathcal{M} \vDash \neg\phi \Leftrightarrow \mathcal{M} \nvDash \phi \Leftrightarrow \mathcal{N} \nvDash \phi \Leftrightarrow \mathcal{N} \vDash \neg\phi$

$\mathcal{M} \vDash \phi \wedge \psi \Leftrightarrow \mathcal{M} \vDash \phi$ and $\mathcal{M} \vDash \psi \Leftrightarrow \mathcal{N} \vDash \phi$ and $\mathcal{N} \vDash \psi \Leftrightarrow \mathcal{N} \vDash \phi \wedge \psi$

if $\mathcal{M} \vDash \exists x\, \phi(x)$ then there is $m \in M$ such that $\mathcal{M} \vDash \phi(m)$ so $\mathcal{N} \vDash \phi(m) \Rightarrow \mathcal{N} \vDash \exists x\, \phi(x)$.

Conversely if $\mathcal{N} \vDash \exists x\, \phi(x)$ then by hypothesis, $\exists m \in M, \mathcal{N} \vDash \phi(m) \Rightarrow \mathcal{M} \vDash \exists x\, \phi(x)$. Hence by induction on $\mathrm{Form}(\mathcal{L})$ we have shown $\mathcal{N} \vDash \mathrm{Diag}_{\mathrm{el}}(\mathcal{M})$, in other words $\mathcal{M} \preccurlyeq \mathcal{N}$. $\square$

For $\phi(x) \in \mathrm{Form}(\mathcal{L})$, we say $n \in \mathcal{N}$ *witnesses* the existential statement $\exists x\, \phi(x)$ if $\mathcal{N} \vDash \exists x\, \phi(x) \Rightarrow \mathcal{N} \vDash \phi(n)$. The Tarski-Vaught test says that any substructure $\mathcal{M} \subseteq \mathcal{N}$ is elementary if and only if whenever there is an existential statement parameterized only by elements of $\mathcal{M}$ that is satisfied by $\mathcal{N}$, it must be witnessed by an element in $\mathcal{M}$.

A direct generalization of the notion of a bijective homomorphism i.e. an *isomorphism* arises: A bijective $\mathcal{L}$-homomorphism is an $\mathcal{L}$-isomorphism. We write $\mathcal{M} \cong \mathcal{N}$ if there is an $\mathcal{L}$-isomorphism between $\mathcal{M}$ and $\mathcal{N}$.

A similar notion is that of *elementary equivalence:* Two $\mathcal{L}$ structures $\mathcal{M}$ and $\mathcal{N}$ are *elementarily equivalent* if $\mathrm{Th}(\mathcal{M}) = \mathrm{Th}(\mathcal{N})$, i.e. for all $\mathcal{L}$-sentences ϕ, $\mathcal{M} \vDash \phi \Leftrightarrow \mathcal{N} \vDash \phi$. We write $\mathcal{M} \equiv \mathcal{N}$. Observe that $\mathcal{M} \preccurlyeq \mathcal{N} \Rightarrow \mathcal{M} \equiv \mathcal{N}$.

Two structures are elementarily equivalent if they are indistinguishable by means of the first-order language $\mathcal{L}$. This is weaker than isomorphism:

Theorem 2.3.[12] *If $\theta : \mathcal{M} \to \mathcal{N}$ is an isomorphism then $\mathcal{M} \equiv \mathcal{N}$.*

Proof. θ is an $\mathcal{L}$-embedding so from proof of Theorem 2.1 we know that for $t \in \mathrm{Term}(\mathcal{L})$, $\theta(t^{\mathcal{M}}) = t^{\mathcal{N}}$.

If an $\mathcal{L}$-formula ϕ is $R(\bar{t})$ for $R \in \mathcal{L}_R$, then we have

$$\mathcal{M} \vDash \phi \Leftrightarrow \bar{t}^{\mathcal{M}} \in R^{\mathcal{M}} \Leftrightarrow \theta(\bar{t}^{\mathcal{M}}) \in \theta(R^{\mathcal{M}}) \Leftrightarrow \bar{t}^{\mathcal{N}} \in R^{\mathcal{N}} \Leftrightarrow \mathcal{N} \vDash \phi.$$

If ϕ is $\neg\psi$ for an $\mathcal{L}$-formula ψ then

$$\mathcal{M} \vDash \phi \Leftrightarrow \mathcal{M} \nvDash \psi \Leftrightarrow \mathcal{N} \nvDash \psi \Leftrightarrow \mathcal{N} \vDash \phi.$$

If ϕ is $\psi \wedge \eta$ then

$$\mathcal{M} \vDash \phi \Leftrightarrow \mathcal{M} \nvDash \psi \text{ and } \mathcal{M} \nvDash \eta \Leftrightarrow \mathcal{N} \nvDash \psi \text{ and } \mathcal{N} \nvDash \eta \Leftrightarrow \mathcal{N} \vDash \phi.$$

If ϕ is $\exists x\, \psi(x)$ then

$$\begin{aligned} \mathcal{M} \vDash \phi &\Leftrightarrow \mathcal{M} \vDash \psi(m) \text{ for some } m \in M \\ &\Leftrightarrow \mathcal{N} \vDash \psi(n) \text{ for some } n \in N \text{ since } \theta \text{ surjective} \\ &\Leftrightarrow \mathcal{N} \vDash \phi \end{aligned}$$

Hence by induction on $\mathrm{Form}(\mathcal{L})$, we have shown that $\mathcal{M} \vDash \phi \Leftrightarrow \mathcal{N} \vDash \phi$ for all $\mathcal{L}$-sentences ϕ, i.e. $\mathcal{M} \equiv \mathcal{N}$. $\square$

Theorem 2.3 says that $\mathrm{Th}(\mathcal{M})$ is an *isomorphism-invariant.*

2.1.1 Theorems of Löwenheim and Skolem

Is it possible to produce more models of a theory from existing ones? Turns out it is:

If $\mathcal{M}$ is an $\mathcal{L}$-structure and S is a set of elements from M, then the *hull* of S, written $\langle S \rangle$ is the smallest substructure of $\mathcal{M}$ containing S.

Lemma 2.4.[10] *If $\mathcal{M}$ is an $\mathcal{L}$-structure and S is a set of elements in M then $|\langle S \rangle| \leq |S| + |\mathcal{L}| + \aleph_0$*

Proof. We define the set S_i inductively for $i < \omega$:

$$S_0 = S \cup \{c^{\mathcal{M}} : c \in \mathcal{L}_C\}$$
$$S_{i+1} = S_i \cup \{f^{\mathcal{M}}(\bar{a}) : f \in \mathcal{L}_F, \bar{a} \in S_i^n\}$$

Set $S_\omega = \bigcup_{i<\omega} S_i$ which is a subset of M. Then $\mathcal{S} = (S_\omega, ...)$ is an $\mathcal{L}$-structure by interpreting $c^{\mathcal{S}} = c^{\mathcal{M}} \in S_\omega$ and relation-symbols and function-symbols as relations and functions of $\mathcal{M}$ over the restricted domain S_ω. Moreover, $\mathcal{S}$ is a substructure of $\mathcal{M}$, and from construction,

$$|\langle S \rangle| \leq |S_\omega| \leq |S| + |\mathcal{L}| + \aleph_0.$$

□

Two important results by Thoralf Skolem and Leopold Löwenheim give a method to build elementary substructures and superstructures of any cardinality from an infinite model of a theory. Ironically, Skolem studied these because he disliked the idea of uncountable structures, and aimed to show *"every countable theory which is satisfiable has a countable model"*. We prove a generalized version of the statement:

Theorem 2.5. ***(Downward Löwenheim-Skolem Theorem)***[2] *For a first-order language $\mathcal{L}$, if an $\mathcal{L}$-theory T has an infinite model $\mathcal{M}$ and κ is an infinite cardinal such that $|\mathcal{L}| \leq \kappa \leq |\mathcal{M}|$ then T has a model with cardinality κ.*

[2] [10] proves this result using Henkin's ideas, but also sketches a shorter, more straightforward proof. For the purposes of this text it suffices to use the latter, the details of which have been filled in.

Proof. Consider any collection S of size κ of elements of $\mathcal{M}$. Define S_i inductively for $i < \omega$:

$$S_0 = \langle S \rangle$$
$$S_{i+1} = \langle\, S_i \cup \{m \in M : \text{for}\, \phi(x) \in \mathrm{Diag}_{\mathrm{el}}(S_i),\ \mathcal{M} \vDash \exists x\, \phi(x) \Rightarrow \mathcal{M} \vDash \phi(m)\} \,\rangle$$

i.e. every existential statement in terms of S_i satisfied by $\mathcal{M}$ is witnessed by some element in S_{i+1}. This gives us a chain $S_0 \subseteq S_1 \subseteq S_2 \subseteq ... \subseteq \mathcal{M}$ of substructures. Let $\mathcal{S} = \bigcup_{i<\omega} S_i$ so that $\mathcal{S} \subseteq \mathcal{M}$. Now by construction, $\mathcal{S}$ satisfies the Tarski-Vaught test and hence from Theorem 2.2, $\mathcal{S} \preccurlyeq \mathcal{M}$. Moreover, $|S| \leq |\mathcal{S}| \leq |\mathcal{L}| + |S| + \aleph_0 = |S|$ so $|\mathcal{S}| = |S| = \kappa$. $\square$

Theorem 2.6. ***(Upward Löwenheim-Skolem Theorem*[3]*)*** *[17] For a first-order language $\mathcal{L}$, if an $\mathcal{L}$-theory T has an infinite model $\mathcal{M}$ and $\kappa \geq |\mathcal{L}|$ is an infinite cardinal, then T has a model with cardinality κ.*

Proof. Define $\mathcal{L}^* = \mathcal{L} \cup \{c_i : i < \kappa\}$ by adding κ constant-symbols to $\mathcal{L}$, and an $\mathcal{L}^*$-theory $T^* = T \cup \{c_i \neq c_j : i, j < \kappa, i \neq j\}$. Every finite subset of T^* is satisfied by $\mathcal{M}$ since $|\mathcal{M}| \geq \aleph_0$, so by compactness theorem T^* is satisfiable. However if $\mathcal{N}' \vDash T^*$ then $|\mathcal{N}'| \geq \kappa$ (since T^* asserts $\mathcal{N}'$ has atleast κ distinct elements.) Use the downward Löwenheim-Skolem theorem to find an elementary substructure $\mathcal{N} \preccurlyeq \mathcal{N}'$ of cardinality κ, and read off $\mathcal{N}$ as an $\mathcal{L}$-structure. Observe $\mathcal{N}' \vDash T \Rightarrow \mathcal{N} \vDash T$ to finish the proof. $\square$

Combined, the theorems assert that first-order languages cannot distinguish between sizes of infinities- for every infinite cardinal bigger than $|\mathcal{L}|$, an $\mathcal{L}$-structure $\mathcal{M}$ has elementary substructures or elementary expansions of that size. This also shows why elementary equivalence is strictly weaker than isomorphism- we can construct two elementarily equivalent structures of different cardinalities. Skölem's motivation was to show that this leads to problems if the theory itself talks about cardinality:

Skölem's 'paradox':[5] First order Zermelo-Fraenkel set theory (ZFC) asserts the existence of uncountable sets. But the downward Löwenheim-Skolem theorem asserts the existence of countable models of ZFC. How can there be a countable model containing uncountable sets? Skölem wanted to use this as an argument against uncountability, but pointed out how it isn't actually a paradox. At best it is an equivocation fallacy: what is 'uncountable'?

[3]Who called it upward Löwenheim-Skolem theorem and not Hÿghenheim-Skolem theorem?

ZFC is written in the language $\{\epsilon\}$ containing one binary relation-symbol. In our (uncountable) model of ZFC, elements are sets (functions are considered to be sets), and the binary relation is interpreted as containment $\in$. *'There are uncountable sets'* then is the first order sentence saying *'there exists an x such that there does not exist a function which is a bijection with domain $\mathbb{N}$ and range x.'* Of course, this is all expressed in terms of ϵ, and the very first problem observed is ϵ might have different interpretations in different models so the interpretation of countability itself changes from model to model. More so, it is possible that there are sets the model 'thinks' are uncountable but are really countable: the bijection between the set and $\mathbb{N}$ is not contained within our domain- and this is possible since there are $2^{\aleph_0}$ functions between a countable set and $\mathbb{N}$ but only $\aleph_0$ of them can be contained within the domain of a countable model. The bijection can be contained in a larger model, of course.

2.1.2 Games of Ehrenfeucht and Fraïssé[10][12]

If the Löwenheim-Skolem theorems give a method to move 'up and down' to establish elementary equivalence, Ehrenfeucht-Fraïssé games give a 'back and forth' method. This was first given by Roland Fraïssé, then formulated as a game by Andrzej Ehrenfeucht. For the rest of the section we assume $\mathcal{L}$ is a first-order language with no function-symbols (this isn't much of a problem since we can simply replace functions with their graphs as relation-symbols). A *partial embedding* between two $\mathcal{L}$-structures $\mathcal{M} = (M, ...)$ and $\mathcal{N} = (N, ...)$ is a function f from $A \subseteq M$ to $B \subseteq N$ such that $f \cup \{(c^{\mathcal{M}}, c^{\mathcal{N}}) : c \in \mathcal{L}_C\}$ preserves the relations and constants of $\mathcal{L}$.

For two $\mathcal{L}$-structures $\mathcal{M} = (M, ...)$ and $\mathcal{N} = (N, ...)$ with disjoint universes, we define a γ-turn two-player game $G_\gamma(\mathcal{M}, \mathcal{N})$. We call the two players Jerry and George. Jerry wants to show $\mathcal{M}$ and $\mathcal{N}$ are 'similar', George wants to stop him. It is played in turns, and the length of the game γ is already decided. At the i^{th} turn, either

George picks an element $m_i \in M$ and challenges Jerry to pick $n_i \in N$, or

George picks an element $n_i \in N$ and challenges Jerry to pick $m_i \in M$.

Jerry wins if $f = \{(m_i, n_i) : i < \gamma\}$ is a partial embedding.

A *strategy* is a function τ such that if George's first k moves are $\bar{a}_{k-1} = (a_0, a_1, ..., a_{k-1})$ then Jerry's k^{th} move is $\tau(\bar{a}_{k-1})$.

τ is a *winning strategy* for Jerry if $\{(a_i, \tau(\bar{a}_i)) : i < \gamma\}$ is a partial embedding, i.e. if playing according to τ helps Jerry win.

For $\gamma < \omega$, we write $\mathcal{M} \sim_\gamma \mathcal{N}$ if Jerry has a winning strategy for $G_\gamma(\mathcal{M}, \mathcal{N})$. Observe that $\sim_\gamma$ is an equivalence relation on the set of $\mathcal{L}$-structures.

Example 2.7. Say Jerry and George play the game for $\mathbb{R}$ and $\mathbb{Q}$ as $\mathcal{L} = \{+, 0\}$-structures. George picks some rational $m_0 \neq 0$ in $\mathbb{R}$, so Jerry is forced to pick some $n_0 \neq 0$ in $\mathbb{Q}$. Now George simply picks an irrational m_1 in $\mathbb{R}$. Whatever n_1 Jerry picks, it will be a rational so there will exist non-zero integers a, b such that $\underbrace{n_0 + n_0 + ... + n_0}_{a \text{ times}} = \underbrace{n_1 + n_1 + ... + n_1}_{b \text{ times}}$ and hence $\{(m_0, n_0), (m_1, n_1)\}$ can never be a partial embedding.
Thus $\mathbb{R} \sim_1 \mathbb{Q}$ but $\mathbb{R} \nsim_2 \mathbb{Q}$.

$G_\omega(\mathcal{M}, \mathcal{N})$ is the (countable) infinite Ehrenfeucht-Fraïssé game, and we say $\mathcal{M}$ is *back and forth equivalent* to $\mathcal{N}$ if Jerry has a winning strategy for $G_\omega(\mathcal{M}, \mathcal{N})$.
We write $\mathcal{M} \sim_\omega \mathcal{N}$ if for all n, $\mathcal{M} \sim_n \mathcal{N}$. Note Jerry having a winning strategy for $G_\omega(\mathcal{M}, \mathcal{N})$ is stronger than $\mathcal{M} \sim_\omega \mathcal{N}$.

Lemma 2.8. *One of the players must have a winning strategy for $G_n(\mathcal{M}, \mathcal{N})$.*[4]

Proof. (Sketch) Let's say Jerry has no winning strategy for $G_n(\mathcal{M}, \mathcal{N})$, i.e. there is a move George can make in round 1 so that nothing Jerry plays can force a win. Say George plays that. Now whatever Jerry does, George still has a move that he can make so that Jerry cannot force a win. And George continues with this- the game has a winning strategy for George.[5] □

We say $\mathcal{M}$ and $\mathcal{N}$ are *n-equivalent* if $\mathcal{M} \vDash \phi \Leftrightarrow \mathcal{N} \vDash \phi$ whenever $\text{depth}(\phi) \leq n$, and write $\mathcal{M} \equiv_n \mathcal{N}$.

Lemma 2.9. *Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathcal{L}$-structures. Then $\mathcal{M} \sim_n \mathcal{N}$ if and only if $\mathcal{M} \equiv_n \mathcal{N}$.*

Proof. We proceed by induction on n:
If $\mathcal{M} \equiv_n \mathcal{N}$ and without loss of generality George plays $a \in M$ on round 1. If $\{\phi_0, \phi_1, ..., \phi_k\}$ is a list of all formulae with depth $< n$ and let $\psi(x)$ be

$$\bigwedge_{\mathcal{M} \vDash \phi_i(a)} \phi_i \wedge \bigwedge_{\mathcal{M} \nvDash \phi_i(a)} \neg \phi_i.$$

[4]See also: Zermelo's theorem, determinacy of closed games

[5]in which case of course, George says *"the equivalence store called, they're running out of you!"*

Then $\text{depth}(\exists x\, \psi(x)) \leq n$ and $\mathcal{M} \models \exists x\, \psi(x)$ so $\mathcal{N} \models \exists x\, \psi(x)$. Thus there is $b \in N$ such that $\mathcal{M} \models \phi(a) \Leftrightarrow \mathcal{N} \models \phi(b)$ whenever $\text{depth}(\phi) < n$. Jerry plays b in round 1. If $n = 1$ Jerry wins and we are done here.
If $n > 1$, add a constant symbol c to $\mathcal{L}$ and get $\mathcal{L}^* = \mathcal{L} \cup \{c\}$. Let $(\mathcal{M}, a)$ and $(\mathcal{N}, b)$ be $\mathcal{L}^*$-structures formed by interpreting c with a and b respectively. We have chosen a and b such that $(\mathcal{M}, a) \equiv_{n-1} (\mathcal{N}, b)$ so by induction hypothesis, Jerry has a winning strategy for $G_{n-1}((\mathcal{M}, a), (\mathcal{N}, b))$. Let this result in a partial $\mathcal{L}^*$-embedding f^*. Then $f = f^* \cup \{(a, b)\}$ is a partial $\mathcal{L}$-embedding between $\mathcal{M}$ and $\mathcal{N}$ so Jerry has a winning strategy for $G_n(\mathcal{M}, \mathcal{N})$.

Conversely if $\mathcal{M} \not\equiv_n \mathcal{N}$, without loss of generality there is some ϕ with $\text{depth}(\phi) < n$ such that $\mathcal{M} \models \exists x\, \phi(x)$ and $\mathcal{N} \models \forall x\, \neg\phi(x)$. George plays $a \in M$ such that $\mathcal{M} \models \phi(a)$. Whatever $b \in N$ Jerry plays, $\mathcal{N} \not\models \phi(b)$. If $n = 1$, George wins and we are done. If $n > 1$, we construct $(\mathcal{M}, a)$ and $(\mathcal{N}, b)$ as above, and $(\mathcal{M}, a) \not\equiv_{n-1} (\mathcal{N}, b)$ so George plays according to his winning strategy for $G_{n-1}((\mathcal{M}, a), (\mathcal{N}, b))$. □

Theorem 2.10. *Let $\mathcal{M}$ and $\mathcal{N}$ be $\mathcal{L}$-structures. Then $\mathcal{M} \sim_\omega \mathcal{N}$ if and only if $\mathcal{M} \equiv \mathcal{N}$.*

Proof. Follows immediately from observing

$$\mathcal{M} \sim_\omega \mathcal{N} \Leftrightarrow \text{ for all } n,\ \mathcal{M} \sim_n \mathcal{N} \Leftrightarrow \text{ for all } n,\ \mathcal{M} \equiv_n \mathcal{N} \Leftrightarrow \mathcal{M} \equiv \mathcal{N}.$$

□

These games are very useful techniques to establish elementary equivalence, and the strategy of forming partial embeddings finds applications elsewhere too. While it is reasonable to believe that isomorphism is 'more fundamental' than elementary equivalence since the latter depends on your language, one may also argue that elementary equivalence is 'more fundamental'- isomorphism relies on the existence of an explicit function between the two structures, and hence on the set theoretic universe we are working in. Similar to the case in Skolem's 'paradox', it might be the case that our set theoretic universe is not large enough to contain the function that establishes isomorphism. Elementary equivalence does not depend on the surrounding universe, and hence free from such problems.

We conclude with a special case where the game is indeed useful in showing isomorphism:

Theorem 2.11. *If $\mathcal{M}$ and $\mathcal{N}$ are countable $\mathcal{L}$-structures, then Jerry has a winning strategy for $G_\omega(\mathcal{M}, \mathcal{N})$ if and only if $\mathcal{M} \cong \mathcal{N}$.*

Proof. If $\mathcal{M} \cong \mathcal{N}$ then Jerry wins by playing according to the isomorphism. Conversely let Jerry have a winning strategy. Then if George plays such that every element in M and N is chosen atleast once, the partial embedding built will be an isomorphism. $\square$

Example 2.12. (This is a question from the *Numbers and Sets* course taught the Cambridge Mathematics Tripos. If you are currently studying Part 1A and haven't seen this question yet I strongly recommend you try solving it on your own before reading the solution below.)

> Find a bijection $f : \mathbb{Q} \to \mathbb{Q} \setminus \{0\}$. Can f be strictly increasing (that is, $f(x) < f(y)$ whenever $x < y$)?[2]

Solution: For the language $\mathcal{L}$ with signature $\{<\}$ (interpreted as the order-relation), we have two $\mathcal{L}$-structures $\mathcal{M} = (\mathbb{Q}, <)$ and $\mathcal{N} = (\mathbb{Q} \setminus \{0\}, <)$. In the Ehrenfeucht-Fraïssé game $G_\omega(\mathcal{M}, \mathcal{N})$, Jerry clearly has a winning strategy since in either set, he can always pick an element between any two elements. Since both the sets are countable, the partial embedding thus build would be an isomorphism, i.e. a strictly increasing bijection.

2.2 Lindström's Classification of First-Order Logic

The sentences you are allowed to write are determined by the 'strength' of your *logic*. We have a set $\mathfrak{S}$, carrying information about the syntax of sentences. A logic $\mathfrak{L}$ is the set $\mathfrak{S}$ with a relation $\vDash_{\mathfrak{L}}$ (the *truth predicate*) between arbitrary structures and elements of $\mathfrak{S}$. We write $\mathfrak{L} = (\mathfrak{S}, \vDash_{\mathfrak{L}})$.[18]

One of the key properties the truth predicate must satisfy[6] is *closure under isomorphism:* for all signatures σ and all $\phi \in \mathfrak{S}$, writing ϕ_σ for the $\mathfrak{L}(\sigma)$-sentence corresponding to ϕ, if $\mathfrak{L}(\sigma)$-structures $\mathcal{M}$ and $\mathcal{N}$ are such that $\mathcal{M} \cong \mathcal{N}$ then $\mathcal{M} \vDash_{\mathfrak{L}} \phi_\sigma$ if and only if $\mathcal{N} \vDash_{\mathfrak{L}} \phi_\sigma$.[18]

If we set $\mathfrak{S}_{\mathfrak{o}}$ to be the set of all first-order sentence-syntaxes, and $\vDash$ as defined in Definition 1.11, we get the *first-order logic* $\mathfrak{L}_{\omega\omega} = (\mathfrak{S}_{\mathfrak{o}}, \vDash)$. The key property is: our existential quantifiers must span over the whole domain

[6] The relation should be closed under isomorphism, negation, conjunction, existential quantification, renaming, and free expansion.

of the structure, and we are allowed finite conjunctions and finitely many quantifiers in our sentences.

What do the two subscripts mean? For cardinals κ and λ, a general *infinitary logic* $\mathfrak{L}_{\kappa\lambda}$ is the logic that allows conjunction of *up to* κ formulae by $\wedge$ or $\vee$, and use of *up to* λ quantifiers in a row. Thus $\mathfrak{L}_{\infty 0}$ allows unrestricted conjunction of quantifier-free formulae, while $\mathfrak{L}_{\omega\infty}$ allows conjunction of finitely many formulae containing unrestricted quantification (such as $\forall(x_i : i \in I)$).[10]

Our quantifiers, however, can only span over elements in the domain. This limits our study mostly to structures algebraic in nature. For example, one can write a highly detailed description of $\mathbb{R}$ as a *field* of real numbers using first order logic, however there is no way to encode the analytic properties of $\mathbb{R}$ since the defining *least upper bound property* begins with quantification over *every subset*. We need stronger some logic for that:

We get the *second-order logic* $\mathfrak{L}^{\mathrm{II}}$ by allowing quantification not only over elements of the domain, but also over relations. This is achieved by adding i-ary relation variables $X^i, Y^i, ...$ to the signature. For instance, $\mathcal{M} \vDash_{\mathfrak{L}^{\mathrm{II}}} \forall X^i\, \phi$ is interpreted as *for every i-ary relation R on the domain of* $\mathcal{M}$, $(\mathcal{M}, R) \vDash_{\mathfrak{L}^{\mathrm{II}}} \phi$.

Then, the least upper bound property can be encoded as:

$$\overbrace{\forall X^1 \quad (((\exists z\, X^1(z))}^{\text{for every non-empty subset}} \;\wedge\; \overbrace{(\exists x\;\, \forall y\, X^1(y) \rightarrow x \geq y))}^{\text{with an upper bound}} \tag{2.1}$$

$$\rightarrow \exists s\, \underbrace{((\forall y\, X^1(y) \rightarrow s \geq y)}_{s \text{ is an upper bound}} \;\wedge\; \underbrace{((\forall x\;\, (\forall y\, X^1(y) \rightarrow x \geq y)) \rightarrow s \leq x))).}_{\text{if } x \text{ is an upper bound then } s \text{ is less than } x} \tag{2.2}$$

2.2.1 Comparing logics[18]

For a signature σ, write $\mathrm{Str}(\sigma)$ to be the set of all σ-structures. For $\phi \in \mathfrak{S}$, ϕ_σ is the σ-sentence determined by ϕ. Write $\mathrm{Mod}_{\mathfrak{L},\sigma}(\phi) = \{\mathcal{M} \in \mathrm{Str}(\sigma) : \mathcal{M} \vDash_{\mathcal{L}} \phi_\sigma\}$.

We say a logic $\mathfrak{L} = (\mathfrak{S}, \vDash_{\mathfrak{L}})$ is *less expressive* than a logic $\mathfrak{L}' = (\mathfrak{S}', \vDash_{\mathfrak{L}'})$, written $\mathfrak{L} \leq \mathfrak{L}'$, if for every $\phi \in \mathfrak{S}$ there is a $\phi' \in \mathfrak{S}'$ such that for every signature σ, $\mathrm{Mod}_{\mathfrak{L},\sigma}(\phi) = \mathrm{Mod}_{\mathfrak{L}',\sigma}(\phi')$. Another way to put this is *every class of structures finitely axiomatizable in* $\mathfrak{L}$ *is finitely axiomatizable in* $\mathfrak{L}'$. Informally, this says everything expressible in $\mathfrak{L}$ is expressible in $\mathfrak{L}'$.

Two logics are *equivalent*, written $\mathfrak{L} \equiv \mathfrak{L}'$, if $\mathfrak{L} \leq \mathfrak{L}'$ and $\mathfrak{L}' \leq \mathfrak{L}$.

2.2.2 Lindström's theorem[18]

We can now discuss what makes first-order logic so powerful. The definitions of compactness property and downward Löwenheim-Skolem property follow immediately from the respective theorems. Then, a slightly weaker[7] version of the theorem is given as:

Theorem 2.13. ***(Lindström's theorem)*** *Let $\mathfrak{L}$ be a logic such that $\mathfrak{L} \geq \mathfrak{L}_{\omega\omega}$. Then $\mathfrak{L}$ has the compactness property and the downward Löwenheim-Skolem property if and only if $\mathfrak{L} \equiv \mathfrak{L}_{\omega\omega}$.*

Proof. (Sketch) Suppose $\mathfrak{L}$ is a logic satisfying both given properties, but some $\phi \in \mathfrak{S}$ is not first-order definable, i.e. $\mathrm{Mod}_{\mathfrak{L},\sigma}(\phi)$ cannot be written as $\mathrm{Mod}_{\mathfrak{L}_{\omega\omega},\sigma}(\psi)$ for any first-order ψ. Assume σ is finite and relational.
From Lemma 1.7 we know only there are only finitely many first order formulae with depth up to n. Call two structures n-equivalent ($\mathcal{M} \equiv_n \mathcal{N}$) if they satisfy the same first-order formulae of depth up to n. Then we have only finitely many n-equivalence classes on $\mathrm{Str}(\sigma)$. Since ϕ is not first-order definable, we can find two σ- structures $\mathcal{M}$ and $\mathcal{N}$ such that:

$$\begin{aligned} &\mathcal{M} \equiv_n \mathcal{N} \\ &\mathcal{M} \vDash_{\mathfrak{L}} \phi \\ &\mathcal{N} \vDash_{\mathfrak{L}} \neg\phi. \end{aligned}$$

Lindström encodes the above in $\psi(n) \in \mathfrak{S}$ using Ehrenfeucht- Fraïssé characterization of n-equivalence, since $\mathcal{M} \equiv_n \mathcal{N} \Rightarrow \mathcal{M} \sim_n \mathcal{N}$ from Lemma 2.9.
By compactness, since this is satisfiable for every n, we have two structures $\mathcal{M}$ and $\mathcal{N}$ such that Jerry has a winning strategy for $G_\omega(\mathcal{M},\mathcal{N})$.[8]
By downward Löwenheim-Skolem property, we can assume $\mathcal{M}$ and $\mathcal{N}$ are countable.
But then by Theorem 2.11, $\mathcal{M} \cong \mathcal{N}$. This is a contradiction, since $\vDash_{\mathfrak{L}}$ is closed under isomorphism (if $\mathcal{M} \cong \mathcal{N}$ then $\mathcal{M} \vDash_{\mathfrak{L}} \phi \Leftrightarrow \mathcal{N} \vDash_{\mathfrak{L}} \phi$). $\square$

This theorem says that any logic more expressive than first-order cannot have both the compactness and the downward Löwenheim-Skolem properties. This is important, for the compactness theorem is one of our most

[7] The stronger version assumes countable compactness instead of compactness.

[8] We use compactness to go from $\mathcal{M} \sim_\omega \mathcal{N}$ to the stronger condition, Jerry having a winning strategy for $G_\omega(\mathcal{M},\mathcal{N})$

useful tools.

An interesting observation would be the result that *any ordered field with the least upper bound property must be isomorphic to* $(\mathbb{R}, <, +, \cdot, 0, 1)$. If the least upper bound property were expressible in first-order (the axioms of an ordered field already are), then by the Löwenheim-Skolem theorems we would have to accomodate models that are non-isomorphic to the real numbers. However the result is true and the least upper bound property cannot be expressed in first order logic.

Chapter 3

Boole, Arrow, and Compactness Revisited

3.1 Boole

The discussion that follows is applicable to any partially ordered set. We shall use a setting of *Boolean Algebras*. Define $\sigma_B = \{\neg, \vee, \wedge, 0, 1\}$ (symbols treated differently from the logical conjunction and negation symbols of the language) and an elementary class axiomatized by:

$$\begin{aligned}
\{\forall x\ x \vee x = x, &\quad \forall x\ x \wedge x = x, \\
\forall xy\ x \vee y = y \vee x, &\quad \forall xy\ x \wedge y = y \wedge x, \\
\forall xy\ (x \wedge y) \vee y = y, &\quad \forall xy\ (x \vee y) \wedge y = y, \\
\forall xyz\ (x \vee y) \vee z = x \vee (y \vee z), &\quad \forall xyz\ (x \wedge y) \wedge z = x \wedge (y \wedge z), \\
\forall xyz\ x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z), &\quad \forall xyz\ x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z), \\
\forall x\ x \vee \neg x = 1, &\quad \forall x\ x \wedge \neg x = 0, \\
0 \neq 1\}
\end{aligned}$$

We write $x \leq y$ if $x \wedge y = x$.[10]

Structures satisfying these axioms are called Boolean Algebras. An example of a Boolean algebra is the *powerset algebra* of a set X: the universe is the powerset of X, '0' and '1' are $\emptyset$ and X respectively, '$\neg$' is set-theoretic complement, '$\wedge$' and '$\vee$' are $\cap$ and $\cup$ respectively, and '$\leq$' is $\subseteq$. We say *Boolean algebra over* X to refer to the powerset algebra of X.

Another example is $\mathbf{2} = \{0, 1\}$, where we refer to the elements 0 and 1 as *false* and *true* respectively. Given any Boolean algebra $\mathcal{B}$, we construct a function $\theta : \mathcal{B} \to \mathbf{2}$ by setting $\theta(a) = \begin{cases} 0 & a \text{ is not nice} \\ 1 & a \text{ is nice} \end{cases}$, for some property called *niceness*. For the discussion to be meaningful, the map will have to be a homomorphism. This immediately imposes some conditions on the elements that can be nice, i.e. elements in $\theta^{-1}(1) = \mathcal{U}$:

$$
\begin{gathered}
0 \notin \mathcal{U}, \quad 1 \in \mathcal{U} \\
\text{if } a \in \mathcal{U} \text{ and } a \leq b \text{ then } b \in \mathcal{U}, \\
\text{if } a, b \in \mathcal{U} \text{ then } a \wedge b \in \mathcal{U}, \\
a \in \mathcal{U} \Leftrightarrow \neg a \notin \mathcal{U}.
\end{gathered}
$$

Any $\mathcal{U} \subseteq \mathcal{B}$ which satisfies the above properties is called an *ultrafilter* on $\mathcal{B}$. Observe that every homomorphism $\mathcal{B} \to \mathbf{2}$ determines a unique ultrafilter from its pre-image of 1. Conversely, every ultrafilter $\mathcal{U}$ on $\mathcal{B}$ determines a unique homomorphism $\theta : \mathcal{B} \to \mathbf{2}$ given by $\theta(a) = \begin{cases} 0 & a \notin \mathcal{U} \\ 1 & a \in \mathcal{U} \end{cases}$.

If only the first three conditions are obeyed, the set is called a *filter.* It is easy to see that every Boolean algebra has atleast one filter- the *trivial filter* $\{1\}$.

Filters and ultrafilters capture the idea of large sets. An ultrafilter on the Boolean algebra on V (an ultrafilter on V) represents the least structure any collection of subsets of V should have to meaningfully classify the subsets while respecting the logical structure (i.e. by a homomorphism onto $\mathbf{2}$). One can look at a filter on V as the collection of sets that contain 'mostly everything'- this gives good intuition on why the definition was chosen:

$\emptyset$ does not contain 'mostly everything' while V does,
if A contains 'mostly everything' and $A \subseteq B$ then B contains 'mostly everything', and
if A and B both contain 'mostly everything' then so does $A \cap B$.

A direct example illustrating this is the *Fréchet filter*, also called the *cofinite filter* on an infinite set V:

$$\{X : V \setminus X \text{ is finite}\}.$$

While this is not an ultrafilter on the powerset algebra of V, it is an ultrafilter on the *finite-cofinite algebra*- the Boolean algebra of all finite subsets of

V and their complements. But can every Boolean algebra have ultrafilters?

A set W of elements of a Boolean algebra $\mathcal{B}$ is said to have the *finite intersection property* if for every finite collection $\Phi \subseteq W$, $\bigwedge \Phi \neq 0$. Clearly, every filter has the finite intersection property. If $W \subseteq \mathcal{B}$ has the finite intersection property, call

$$\bigcap\{\mathscr{F} : \mathscr{F} \text{ is a filter over } \mathcal{B},\ W \subseteq \mathscr{F}\}$$

the *filter generated by* W. Observe that this indeed is a filter.

Theorem 3.1. ***(Maximality of ultrafilters)***[7] *For a Boolean algebra $\mathcal{B}$, the following are equivalent:*

(i) $\mathscr{U}$ is an ultrafilter on $\mathcal{B}$.

(ii) $\mathscr{U}$ is a maximal filter on $\mathcal{B}$: no filter $\mathscr{F}$ on $\mathcal{B}$ satisfies $\mathscr{U} \subsetneq \mathscr{F}$.

Proof. $(i) \Rightarrow (ii)$: Assume (i) and let $\mathscr{F}$ be a filter on $\mathcal{B}$ such that $\mathscr{U} \subsetneq \mathscr{F}$. If $a \in \mathscr{F} \setminus \mathscr{U}$ then $a \notin \mathscr{U} \Rightarrow \neg a \in \mathscr{U} \Rightarrow \neg a \in \mathscr{F}$. But then $0 = a \wedge \neg a \in \mathscr{F}$, a contradiction since $\mathscr{F}$ is a filter.
$(ii) \Rightarrow (i)$: If $\mathscr{U}$ is a maximal filter, $0 \notin \mathscr{U}$ hence for any $a \in \mathcal{B}$ at most one of a or $\neg a$ can be in $\mathscr{U}$. Suppose for some $a \in \mathcal{B}$, $\neg a \notin \mathscr{U}$.
Let $\mathscr{V} = \mathscr{U} \cup \{a\}$. $\mathscr{V}$ has the finite intersection property: for any finite collection $\Phi \subseteq \mathscr{V}$, $\bigwedge \Phi$ is of the form x or $x \wedge a$ for some $x \in \mathscr{U}$. Consequently, $x \neq 0$ and since $\neg a \notin \mathscr{U}$, we cannot have $x \leq \neg a$, i.e. $x \wedge a \neq 0$.
We can define $\mathscr{F}$ to be the filter generated by $\mathscr{V}$. Now $\mathscr{F}$ itself is a filter over $\mathcal{B}$ and hence $\mathscr{U}$ is not a proper subset of $\mathscr{F}$ by maximality of $\mathscr{U}$. However, from definition of $\mathscr{F}$, we have $\mathscr{U} \subseteq \mathscr{V} \subseteq \mathscr{F}$, thus $\mathscr{U} = \mathscr{V} = \mathscr{F}$ and $a \in \mathscr{U}$. □

Theorem 3.2. ***(Boolean prime ideal theorem)***[7] *For a Boolean algebra $\mathcal{B}$, let $W \subseteq \mathcal{B}$ have the finite intersection property. Then there is an ultrafilter $\mathscr{U}$ on $\mathcal{B}$ such that $W \subseteq \mathscr{U}$.*

Proof. Let $X = \{\mathscr{F} : \mathscr{F} \text{ is a filter over } \mathcal{B},\ W \subseteq \mathscr{F}\}$. The filter generated by W is in X so the set is non-empty. For any non-empty chain C of filters in X, $\bigcup C$ itself lies in X. Then by Zorn's lemma, X has a maximal element say $\mathscr{U}$. Claim that $\mathscr{U}$ is a maximal filter on $\mathcal{B}$: if a filter $\mathscr{U}'$ is such that $\mathscr{U} \subseteq \mathscr{U}'$ then $W \subseteq \mathscr{U}'$, i.e. $\mathscr{U}' \in X$. But $\mathscr{U}$ is maximal in X, so $\mathscr{U}' = \mathscr{U}$. Then, from Theorem 3.1, it follows that $\mathscr{U}$ is an ultrafilter. □

The Boolean prime ideal theorem cannot be proven without invoking the axiom of choice in some form. In fact, it is a strictly[1] weaker version of the axiom of choice- hence becoming a kind of set theoretic choice principle. It gets its name from the fact that it is originally stated as a theorem about *ideals*- the dual notion of filters. Similarly, the dual of an ultrafilter is a *prime ideal*. The prime ideal corresponding to the ultrafilter $\mathscr{U}$ is the set $\{a : \neg a \in \mathscr{U}\}$. While the ultrafilter corresponds to the pre-image of 1 in the homomorphism $\mathcal{B} \to \mathbf{2}$, the prime ideal is the pre-image of 0. The Boolean prime ideal theorem then says that there are enough prime ideals on the Boolean algebra to extend every ideal to a maximal one, or in terms of duals, to extend every filter to an ultrafilter. It can be shown that this result is equivalent to the Compactness theorem![10]

Call an ultrafilter *principal* if it is of the form $\{x : a \leq x\}$ for some $a \in \mathcal{B}$, and *non-principal* otherwise. A simple corollary of the Boolean prime ideal theorem is that non-principal ultrafilters exist on infinite sets, as can be shown by extending the cofinite filter on the set to an ultrafilter on the powerset algebra.

Theorem 3.3. *A principal filter $\mathscr{U}$ on a set V must be of the form $\{X \subseteq V : a \in X\}$ for some $a \in V$.*

Proof. Since $\mathscr{U}$ is non-empty, by principality it has some smallest (under the order defined by $<$) non-empty element A. Hence there is an $a \in A$ and $A \setminus \{a\} \notin \mathscr{U}$ which means $(V \setminus (A \setminus \{a\})) \in \mathscr{U}$. But then $(V \setminus (A \setminus \{a\})) \cap A = \{a\} \in \mathscr{U}$, so $\mathscr{U}' = \{X \subseteq V : \{a\} \in X\} \subseteq \mathscr{U}$. Observe $\mathscr{U}'$ itself is an ultrafilter on V, so by maximality of ultrafilters we must have $\mathscr{U} = \mathscr{U}'$. $\square$

Theorem 3.4. *An ultrafilter $\mathscr{U}$ on a finite set V must be principal.*

Proof. Since V is finite and $\mathscr{U}$ is non-empty, by well-ordering principle[2] it has some element A with the lowest cardinality. A must be non-empty. Hence there is an $a \in A$ and $A \setminus \{a\} \notin \mathscr{U}$ which means $(V \setminus (A \setminus \{a\})) \in \mathscr{U}$. But then $(V \setminus (A \setminus \{a\})) \cap A = \{a\} \in \mathscr{U}$, so $\mathscr{U}' = \{X \subseteq V : \{a\} \in X\} \subseteq \mathscr{U}$. Observe $\mathscr{U}'$ itself is a principal ultrafilter on V, so by maximality of ultrafilters we must have $\mathscr{U} = \mathscr{U}'$. $\square$

The result that there are no non-principal ultrafilters on a finite set has interesting consequences in Voting theory.

[1] See Halpern and Lévy [1971]

[2] compare with proof of Theorem 3.3

3.2 Arrow's Theorem[3][15][16][19]

Let V be a discrete set (the set of voters[3]), the elements of which have to order a finite set S with atleast 3 elements (the set of candidates). If $O(S)$ is the space of all total orderings on S, we define a *voting system* to be a triple (V, S, F) where $F : O(S)^V \to O(S)$ is a function that takes the preference orders of all voters as input (in the form of a function $f : i \mapsto <_i$ called the *preference profile*) and outputs *the result* $<_f \in O(S)$.
For $a, b \in S$ and $X \subseteq V$, write $a <_X b$ if $a <_i b$ for all $i \in X$.

Definition 3.5. A voting system (V, S, F) is *perfect* if for all $a, b \in S$ it obeys

(i) *Consensus*: for a preference profile $f : i \mapsto <_i$, if $a <_V b$ then $a <_f b$.

(ii) *Independence of irrelevant alternatives*: for two preference profiles $f : i \mapsto <_i$ and $f' : i \mapsto <'_i$, if for all i, $a <_i b \Leftrightarrow a <'_i b$, then $a <_f b \Leftrightarrow a <_{f'} b$.

(iii) *Non-dictatorship*: there is no $d \in V$ such that $a <_d b \Rightarrow a <_f b$.

We discussed in the previous section how filters capture the idea of 'large subsets'. In fact, perfect voting systems are closely related with ultrafilters:

Theorem 3.6. *If $\mathscr{U}$ is a non-principal ultrafilter on a set V, then there is a unique voting system (V, S, F) such that for a preference profile $f : i \mapsto <_i$ and distinct $a, b \in S$, $a <_f b$ if and only if $a <_X b$ for some $X \in \mathscr{U}$. Moreover, the voting system is perfect.*

Proof. Define F as given. Knowing pairwise relative orderings uniquely determines the resulting ordering on S: for $a, b, c \in S$, if $a >_f b$ and $b >_f c$ then we must have $a >_A b$ and $b >_B c$ for some $A, B \in \mathscr{U}$. But then $a >_{A \cap B} c$ and $A \cap B \in \mathscr{U}$, so the resulting order $a >_f b >_f c$ is determined. So it remains to show that the pairwise orderings are determined uniquely: say $a, b \in S$, and $X = \{i \in V : a <_i b\}$ be the set of people who preferred b to a. Then naturally, $V \setminus X = \{i \in V : a >_i b\}$ is the set of people who prefer a to b. Now $\mathscr{U}$ is an ultrafilter, so exactly one of X or $V \setminus X$ is in $\mathscr{U}$, without loss of generality $X \in \mathscr{U}$. Then $a <_X b$, so $a <_f b$, i.e. a result exists in all cases. Also, every set Y such that $a >_Y b$ must be a subset of $V \setminus X$ so $Y \notin \mathscr{U}$, and F cannot return both $a <_f b$ and $a >_f b$.
Consensus holds, since V is in $\mathscr{U}$ for every ultrafilter $\mathscr{U}$. The system is also

[3]Not necessarily finite.

independent of irrelevant alternatives, since the only factor in deciding the relative ordering of two elements was their relative ordering in the voting profile. Non-dictatorship holds from non-principality of $\mathcal{U}$.
Thus the voting system is perfect. $\square$

The converse is true as well. In a voting system (V, S, F), call a subset $W \subseteq V$ a *winning coalition* if for $a, b \in S$ we have $a <_f b$ whenever $a <_i b$ for all $i \in W$. This means that a winning coalition can determine the final ordering of any $a, b \in S$ against the strongest opposition.

Theorem 3.7. *The set $\mathcal{U}$ of winning coalitions in any perfect voting system (V, S, F) forms a non-principal ultrafilter on V.*

Proof. Clearly, $\emptyset \notin \mathcal{U}$ and $V \in \mathcal{U}$ from consensus.
If A is a winning coalition then $B \supseteq A$ is automatically a winning coalition.
If A and B are winning coalitions, A and B cannot be disjoint. Then for $a, b, c \in S$ consider the following voting profile:

$$f : \begin{cases} a >_i b >_i c & i \in A \setminus B \\ b >_i c >_i a & i \in B \setminus A \\ b >_i a >_i c & i \in V \setminus (A \cup B) \\ c >_i a >_i b & i \in A \cap B \end{cases}$$

Then $a >_A b$ and $c >_B a$. Since the two are winning coalitions, the result must be $c >_f a >_f b$. From independence of irrelevant alternatives, this shows that $A \cap B$ must be a winning coalition since it is the only set asserting $c > b$.
If A is not a winning coalition, then for the voting profile

$$f : \begin{cases} a <_i b & i \in A \\ a >_i b & i \in V \setminus A \end{cases}$$

the result is $a >_f b$. But this is exactly the condition for $V \setminus A$ to be a winning coalition.
Thus the set of winning coalitions is an ultrafilter on V. If it were principal, from Theorem 3.3 the voting system would be a dictatorship, hence the ultrafilter must be non-principal. $\square$

Corollary 3.8. ***(Arrow's Theorem)*** *Any voting system (V, S, F) with a finite number of voters cannot be perfect.*

Proof. Follows immediately from Theorem 3.4. $\square$

3.3 Ramsey's results[4][10]

Let X be a set, linearly ordered by $<$. Then for a positive integer k, we write $[X]^k$ to denote the set of all strictly increasing k-tuples in X. For a function f with domain $[X]^k$, we say $Y \subseteq X$ is *f-indiscernible* if for all $\bar{a}, \bar{b} \in [Y]^k$, $f(\bar{a}) = f(\bar{b})$ i.e. f cannot distinguish between subsets of k elements in $[Y]$. Similarly if an $\mathcal{L}$-structure $\mathcal{X} = (X, ...)$ is linearly ordered by $<$ and Φ is a collection of $\mathcal{L}$-formulae in k free variables, we say a subset $Y \subseteq X$ is Φ-*indiscernible* if for every $\phi \in \Phi$ and $\bar{a}, \bar{b} \in [Y]^k$, $\mathcal{X} \vDash \phi(\bar{a}) \leftrightarrow \phi(\bar{b})$.

Example 3.9. f X is a vector space with some linear ordering on it, the set Y of bases is ϕ-indiscernible for every first-order formula ϕ, since for any two strictly increasing k-tuples from Y there is an automorphism of X which takes one to another.

Frank Ramsey's work focussed on establishing combinatorial results about the existence of indiscernible subsets in large sets. This forms an entire branch of combinatorics called Ramsey theory, and the central idea is that if one adds 'enough' disorder to the system it is inevitable to create some pattern. We first look at notation, introduced by Paul Erdös and Richard Rado.

Let λ, μ and ν be cardinals and k a positive integer. We write

$$\lambda \to (\mu)^k_\nu$$

to mean if X is any linearly ordered set of cardinality λ, and $f : [X]^k \to \nu$ is a function then X has an f-indiscernible subset of cardinality μ. Stated in terms of the more familiar 'graph-colourings', the notation says that in a graph of λ points, if each subset of k points is assigned one of ν colours then there will be atleast μ subsets all of the same colour.
Observe that if $\lambda' \geq \lambda$, $\mu' \leq \mu$, $\nu' \leq \nu$ and $k' \leq k$ then $\lambda' \to (\mu')^{k'}_{\nu'}$ whenever $\lambda \to (\mu)^k_\nu$.

Example 3.10. For all positive integers n, $n \to (2)^1_{n-1}$, i.e. when n points are coloured in $n-1$ colours then you can always find 2 of the same colour. This result has a name: *the pigeonhole principle.* It also has an infinite version: for all positive integers n, $\omega \to (\omega)^1_n$ i.e. if an infinite set is partitioned into finitely many parts, atleast one will be infinite.[4][5] In fact, something

[4]Similar arguments can be made with countability: $\omega_1 \to (\omega_1)^1_\omega$, i.e. if an uncountable set is partitioned into countably many pieces then atleast one must be uncountable.

[5]In fact, non-principal ultrafilters are a form of infinitary pigeonhole principle: if an infinite set is partitioned into finitely many parts, exactly one will be in the ultrafilter.

stronger is true:

Theorem 3.11. ***(Ramsey's theorem, infinite form)*** *For all positive integers k and n, we have $\omega \to (\omega)^k_n$.*

Proof. We first prove for $n = 2$.
Let $\mathbb{N} = \{0, 1, 2, ...\}$ be the set of cardinality ω, and we have a 2-colouring of $[\mathbb{N}]^k$ given by $[\mathbb{N}]^k = A \cup B$. Given a non-principal ultrafilter $\mathcal{U}$, we inductively define 2-colourings of $[\mathbb{N}]^i$ for $i \leq k$:
Let A^i be the set of all red elements of $[\mathbb{N}]^i$, and B^i be the set of all blue elements. Then let $A^k = A$ and $B^k = B$. If $[\mathbb{N}]^i = A^i \cup B^i$ then we say any element P of $[\mathbb{N}]^{i-1}$ is red if it is in 'mostly red i-sets' i.e. the set $X(P) = \{a \in \mathbb{N} : a > \max(P), P \cup \{a\} \text{ is red}\}$ is an element of the ultrafilter $\mathcal{U}$.

$$A^{i-1} = \{P \in [\mathbb{N}]^{i-1} : X(P) = \{a : a > \max(P), P \cup \{a\} \in A^i\} \in \mathcal{U}\},$$
$$B^{i-1} = \{P \in [\mathbb{N}]^{i-1} : X(P) = \{a : a > \max(P), P \cup \{a\} \in B^i\} \in \mathcal{U}\}.$$

Observe from the properties of $\mathcal{U}$ it is true that A^{i-1} and B^{i-1} form a 2-colouring of $[\mathbb{N}]^i$. Thus we get a 2-colouring of $\mathbb{N}$ by A^1 and B^1. Since $A^1 = (B^1)^c$, without loss of generality, $A^1 \in \mathcal{U}$.
We inductively define a sequence $j_0 < j_1 < ... < j_m < ...$: choose $j_0 \in A^1$. If $\{j_0, j_1, ..., j_m\}$ is such that all its subsets are red, then for $P \subseteq \{j_0, j_1, ..., j_m\}$, the set $X(P)$ is in $\mathcal{U}$. Since there can only be finitely many such subsets,

$$Y = \bigcap_{P \subseteq \{j_0, j_1, ..., j_m\}} X(P) \in \mathcal{U}.$$

Since the ultrafilter is non-principal, Y must be infinite so we can choose $j_{m+1} > j_m$ in Y. Then every subset of $\{j_0, j_1, ..., j_{m+1}\}$ is red. This lets us build an infinite set $J = \{j_0, j_1, ...\} \subseteq \mathbb{N}$ such that every finite subset of J of size $\leq k$ is red, hence $[J]^k \subseteq A$. Since every set of size ω can be bijectively mapped to $\mathbb{N}$, we have $\omega \to (\omega)^k_2$ for every positive integer k.
Now given a positive integer n, a set I of cardinality ω, and an n- colouring of $[I]^k$ given by sets $A_0, A_1, ..., A_{n-1}$ we must have an infinite set $I_0 \subseteq I$ such that $[I_0]^k \subseteq A_0$ or $[I_0]^k \subseteq A_1 \cup A_2 \cup ... \cup A_{n-1}$. If $[I_0]^k \subseteq A_0$ then done, else we can similarly find $I_1 \subseteq I_0$ such that $[I_1]^k \subseteq A_1$ or $[I_1]^k \subseteq A_2 \cup A_3 \cup ... \cup A_{n-1}$. In this way there must be an infinite subset $J \subseteq I$ such that $[J]^k$ is contained in one of the n partitions, thus proving $\omega \to (\omega)^k_n$. □

Stated for $k = 2$, this says for any n-colouring on a complete infinite graph, one can always find a *complete monochromatic infinite graph*, i.e. an infinite collection of points such that they are all connected to each other with edges of the same colour.

Corollary 3.12. ***(Ramsey's theorem, finite form)*** *For positive integers m, n and k there exists a positive integer l such that $l \to (m)^k_n$.*

Proof. Clearly, must have $l \geq n$. Assume for the sake of contradiction that there is no $l < \omega$ such that $l \to (m)^k_n$. Consider the signature $\mathcal{L} = \{<, f, 1, ..., n-1, n\}$ where $<$ is a binary relation-symbol, f is an k-ary function-symbol and $1, ..., n$ are constant-symbols. Let ϕ be the sentence

$$\exists x_1 \exists x_2 ... \exists x_m, \left(\bigwedge_{1 \leq i < j \leq m} (x_i \neq x_j) \right) \wedge$$
$$\left(\bigvee_{i=1,2,...,n} \left(\bigwedge_{y_1, y_2, ..., y_k \in \{x_1, x_2, ..., x_m\}} ((y_1 < y_2 < ... < y_k) \wedge f(y_1, y_2, ..., y_k) = i) \right) \right).$$

ϕ says there is a subset of size $\geq m$ which is f-indiscernible. Let $\mathcal{A}_l$ be an $\mathcal{L}$-structure of cardinality $l (\geq n)$ in which all constant-symbols have distinct interpretations. Since no $l < \omega$ satisfies $l \to (m)^k_n$, all $\mathcal{A}_l$ satisfy $\neg\phi$. Since $\neg\phi$ has arbitrarily large finite models, by compactness it has an infinite model. But this contradicts Theorem 3.11. $\square$

These numbers l are called *Ramsey numbers*. The standard example is $6 \to (3)^2_2$, i.e. in a 2-colouring of the complete graph on 6 points, one can always find a monochromatic triangle. The precise value 6 is found by a neat pigeonhole argument which can be found in any introduction to the subject. Indeed, Corollary 3.12 only proves the existence of such numbers, and gives no estimate on their size. Paul Erdös sums up the struggle in an interesting quote-

> Suppose aliens invade the earth and threaten to obliterate it in a year's time unless human beings can find the Ramsey number for red five and blue five. We could marshal the world's best minds and fastest computers, and within a year we could probably calculate the value. If the aliens demanded the Ramsey number for red six and blue six, however, we would have no choice but to launch a preemptive attack.

3.4 The Ultraproduct Construction

The Löwenheim-Skolem theorem constructs elementary extensions of models by adding enough constant-symbols to the vocabulary. There is a more straightforward method of extending models, based on the recurring mathematical theme of taking *direct products*.

Suppose we have a set A_i for each i in a non-empty set I. Then we define the direct product $\prod_I A_i$ to be the set of all functions $a : I \to \bigcup_I A_i$ such that $a(i) \in A_i$.

Definition 3.13. For a signature $\mathcal{L}$ and $\mathcal{L}$-structures $\mathcal{A}_i = (A_i, ...)$ for each i in a non-empty set I, the direct product $\prod_I \mathcal{A}_i$ is the $\mathcal{L}$-structure $\mathcal{B}$ such that

(i) the domain of $\mathcal{B}$ is $B = \prod_I A_i$,

(ii) for each constant-symbol $c \in \mathcal{L}$, $c^{\mathcal{B}} = a \in B$ such that $a(i) = c^{\mathcal{A}_i}$,

(iii) for each n-ary function-symbol $f \in \mathcal{L}$ and $\bar{a} = (a_0, a_1, ..., a_{n-1}) \in B^n$, $f^{\mathcal{B}}(\bar{a}) = b \in B$ such that $b(i) = f^{\mathcal{A}_i}(\bar{a}(i))$,

(iv) for each n-ary relation-symbol $R \in \mathcal{L}$ and $\bar{a} \in B^n$, $\bar{a} \in R^{\mathcal{B}}$ if and only if for every $i \in I$, $\bar{a}(i) \in R^{\mathcal{A}_i}$.

Note when we write $\bar{a}(i)$ we always mean $(a_0(i), a_1(i), ..., a_{n-1}(i))$ and never a_i.

This definition immediately brings with it the canonical *projection maps* $p_i : \prod_I \mathcal{A}_i \to A_i$ given by $p_i(a) = a(i)$ which exist for each $i \in I$. These are homomorphisms, as can be easily verified. By setting all $\mathcal{A}_i = \mathcal{A}$ for some fixed $\mathcal{L}$-structure $\mathcal{A}$, we determine *the I^{th} power of $\mathcal{A}$*, written $\mathcal{A}^I = \prod_I \mathcal{A}_i$. Then we define the *diagonal embedding* to be the map $d : \mathcal{A} \to \mathcal{A}^I$ such that for $a \in \mathcal{A}$, $d(a)$ is the constant map with value a.[10]

3.4.1 Reduced Products[10]

Suppose we have a set A_i for each i in a non-empty set I. Given a filter $\mathscr{F}$ on I, we can define a relation $\sim$ on $\prod_I A_i$ given by

$$a \sim b \Leftrightarrow \{i : a(i) = b(i)\} \in \mathscr{F}.$$

Proposition 3.14. *$\sim$ is an equivalence relation.*

Proof. $I \in \mathscr{F}$, so for all $a \in \prod_I A_i$, $a \sim a$ and $\sim$ is reflexive.
The definition is symmetric in a and b, so $\sim$ is symmetric.
If $a \sim b$ and $b \sim c$, then $\{i : a(i) = b(i)\} \cap \{i : b(i) = c(i)\} \subseteq \{i : a(i) = c(i)\}$. Since filters are closed under finite intersections and superset, we have $a \sim c$, so $\sim$ is transitive. $\square$

Write $a/\mathscr{F}$ for the equivalence class of a- this is the set of functions which agree with a 'mostly everywhere[6]'. We define the *reduced product* $\prod_I A_i/\mathscr{F}$ to be the set of all $\sim$-equivalence classes.

This definition extends to reduced products of structures: for signature $\mathcal{L}$, let $\mathcal{A}_i = (A_i, ...)$ be $\mathcal{L}$-structures for each i in a non-empty set I, and $\mathcal{B} = \prod_I \mathcal{A}_i$. If $\mathscr{F}$ is a filter on I, we define the reduced product $\prod_I \mathcal{A}_i/\mathscr{F}$ to be the $\mathcal{L}$-structure $\mathcal{C}$ such that

(i) the universe of $\mathcal{C}$ is $C = \prod_I A_i/\mathscr{F}$,

(ii) for constant-symbol $c \in \mathcal{L}$, $c^{\mathcal{C}} = c^{\mathcal{B}}/\mathscr{F}$,

(iii) for n-ary function-symbol $f \in \mathcal{L}$ and $\bar{a} = (a_0, a_1, ..., a_{n-1}) \in B^n$, $f^{\mathcal{C}}(\bar{a}/\mathscr{F}) = f^{\mathcal{B}}(\bar{a})/\mathscr{F}$,

(iv) for n-ary relation-symbol $R \in \mathcal{L}$ and $\bar{a} \in B^n$, $\bar{a}/\mathscr{F} \in R^{\mathcal{C}}$ if and only if $\{i \in I : \bar{a}(i) \in R^{\mathcal{A}_i}\} \in \mathscr{F}$.

Note that we write $\bar{a}/\mathscr{F}$ to mean $(a_0/\mathscr{F}, a_1/\mathscr{F}, ..., a_{n-1}/\mathscr{F})$. Use properties of filters to make sure that the interpretation of relation-symbols is sound. Observe then that the direct product $\prod_I \mathcal{A}_i$ is the reduced product $\prod_I \mathcal{A}_i/\{I\}$.

The *reduced power* $\mathcal{A}^I/\mathscr{F}$ is obtained by letting all $\mathcal{A}_i = \mathcal{A}$ for a fixed structure $\mathcal{A}$, and the diagonal embedding $e : \mathcal{A} \to \mathcal{A}^I/\mathscr{F}$ is given by $e(a) = b/\mathscr{F}$ where $b(i) = a$ for all $i \in I$.

Example 3.15. (This is a question from the *Numbers and Sets* course taught the Cambridge Mathematics Tripos. If you are currently studying Part 1A and haven't seen this question yet I strongly recommend you try solving it on your own before reading the solution below.)

> Each of an infinite set of Trappist set theorists is going to a party, where each will receive a coloured hat, either red or blue. Each

[6] See discussion on filters

person will be able to see every hat but his own. After all hats are assigned, the set theorists must, simultaneously, each write down (in silence, obviously) a guess as to their own hat colour. You are asked to supply them with a strategy such that, should they follow it, only finitely many of them will guess wrongly. Can you?[1]

Solution: There is a strategy: let $C = \{\text{red, blue}\}$ be the set of hat-colours. If I is the set of all set theorists and $\mathscr{F}$ is the Fréchet filter on I, the set theorists all memorize a representative element[7] from each equivalence class in $C^I/\mathscr{F}$. Since they can observe all hats but one, it is possible to place the sequence of hat-colours in an equivalence class $a/\mathscr{F}$ represented by a. The set theorists then guess whichever hat-colour a assigns to them.

If the reduction is by an ultrafilter, the reduced product has nice properties.

3.4.2 Ultraproducts

An *ultraproduct* is a reduced product $\prod_I \mathcal{A}_i/\mathscr{U}$ where $\mathscr{U}$ is an ultrafilter on the set I. An *ultrapower* is a reduced power $\mathcal{A}^I/\mathscr{U}$.

Theorem 3.16. ***(Łoś's[8] Theorem)***[7] *For a first-order language $\mathcal{L}$, let $\{\mathcal{A}_i : i \in I\}$ be a collection of $\mathcal{L}$-structures, where I is a non-empty set with an ultrafilter $\mathscr{U}$ on it. Then for any $\mathcal{L}$-formula $\phi(\bar{x})$ and every tuple $\bar{a} \in \prod_I \mathcal{A}_i$, $\prod_I A_i/\mathscr{U} \vDash \phi(\bar{a}/\mathscr{U})$ if and only if $\{i : \mathcal{A}_i \vDash \phi(\bar{a}(i))\} \in \mathscr{U}$.*

Proof. We proceed by induction on the set of $\mathcal{L}$-formulae.
If ϕ is atomic, then the statement is true from the way relation-symbols are interpreted in reduced products.
If ϕ is $\neg\psi$ then

$$\begin{aligned}\prod_I A_i/\mathscr{U} \vDash \phi(\bar{a}/\mathscr{U}) &\Leftrightarrow \prod_I A_i/\mathscr{U} \nvDash \psi(\bar{a}/\mathscr{U})\\ &\Leftrightarrow \{i : \mathcal{A}_i \vDash \psi(\bar{a}(i))\} \notin \mathscr{U}\\ &\Leftrightarrow \{i : \mathcal{A}_i \nvDash \psi(\bar{a}(i))\} \in \mathscr{U}\\ &\Leftrightarrow \{i : \mathcal{A}_i \vDash \phi(\bar{a}(i))\} \in \mathscr{U}\end{aligned}$$

[7]Using the axiom of choice
[8]Pronounced *wash*

If ϕ is $\psi \wedge \theta$ then

$$\begin{aligned}
\prod_I A_i/\mathcal{U} \vDash \phi(\bar{a}/\mathcal{U}) &\Leftrightarrow \prod_I A_i/\mathcal{U} \vDash \psi(\bar{a}/\mathcal{U}) \wedge \theta(\bar{a}/\mathcal{U}) \\
&\Leftrightarrow \prod_I A_i/\mathcal{U} \vDash \psi(\bar{a}/\mathcal{U}) \text{ and } \prod_I A_i/\mathcal{U} \vDash \theta(\bar{a}/\mathcal{U}) \\
&\Leftrightarrow \{i : \mathcal{A}_i \vDash \psi(\bar{a}(i))\} \in \mathcal{U} \text{ and } \{i : \mathcal{A}_i \vDash \theta(\bar{a}(i))\} \in \mathcal{U} \\
&\Leftrightarrow \{i : \mathcal{A}_i \vDash \psi(\bar{a}(i))\} \cap \{i : \mathcal{A}_i \vDash \theta(\bar{a}(i))\} \in \mathcal{U} \\
&\Leftrightarrow \{i : \mathcal{A}_i \vDash \psi(\bar{a}(i)) \wedge \theta(\bar{a}(i))\} \in \mathcal{U} \\
&\Leftrightarrow \{i : \mathcal{A}_i \vDash \phi(\bar{a}(i))\} \in \mathcal{U}
\end{aligned}$$

If $\phi(\bar{x})$ is $\forall \bar{y}\ \psi(\bar{y}, \bar{x})$ then (assume without loss of generality that the n-tuple $\bar{y}$ is free in ψ, else it would fall into one of previous categories.)

$$\begin{aligned}
\prod_I \mathcal{A}_i/\mathcal{U} \vDash \phi(\bar{a}/\mathcal{U}) &\Leftrightarrow \prod_I \mathcal{A}_i/\mathcal{U} \vDash \forall \bar{y}\ \psi(\bar{y}, \bar{a}/\mathcal{U}) \\
&\Leftrightarrow \text{for all } \bar{b} \in (\prod_I A_i)^n,\ \prod_I \mathcal{A}_i/\mathcal{U} \vDash \psi(\bar{b}/\mathcal{U}, \bar{a}/\mathcal{U}) \\
&\Leftrightarrow \text{for all } \bar{b} \in (\prod_I A_i)^n,\ \{i : \mathcal{A}_i \vDash \psi(\bar{b}(i), \bar{a}(i))\} \in \mathcal{U} \\
&\Leftrightarrow \{i : \text{for all } \bar{b}(i) \in (A_i)^n,\ \mathcal{A}_i \vDash \psi(\bar{b}(i), \bar{a}(i))\} \in \mathcal{U} \\
&\Leftrightarrow \{i : \mathcal{A}_i \vDash \forall \bar{y}\ \psi(\bar{y}, \bar{a}(i))\} \in \mathcal{U} \\
&\Leftrightarrow \{i : \mathcal{A}_i \vDash \phi(\bar{a}(i))\} \in \mathcal{U}
\end{aligned}$$

Thus by induction on the set Form($\mathcal{L}$), the statement is true. □

Corollary 3.17. ***(Transfer principle)***[13] *For a language $\mathcal{L}$, let $\mathcal{M} = (M, ...)$ be an $\mathcal{L}$-structure. If I is a set with an ultrafilter $\mathcal{U}$ on it. Then for every $\mathcal{L}$-sentence ϕ, $\mathcal{M}^* = (M^I/\mathcal{U}, ...) \vDash \phi$ if and only if $\mathcal{M} \vDash \phi$. In other words, Any structure is elementarily equivalent to its ultrapowers.*[7]

Łoś's is important enough to be dubbed *the fundamental theorem of ultraproducts.* A sentence is true in the ultraproduct if and only if it is true in 'a large subset' of structures- one can imagine the structures 'voting' if the sentence is true or not. But why ultraproducts? Most of model-theoretic results can be derived without using ultraproducts, as we have so far. In fact [12] barely mentions ultraproducts in his extended discussion of model theory. Historically, however the theorem is important because of its extensive application in set theory and non-standard analysis. The construction is also generally applicable in a variety of situations (not just models) and a neat trick to have up your sleeve.[9] Also, it provides a characterization of elementary classes.

Corollary 3.18. ***(Compactness theorem)***[7] *Let Σ be a set of $\mathcal{L}$- sentences and $I = \{X : X \subseteq \Sigma,\ X \text{ is finite}\}$. If for all $i \in I$ there exists a model $\mathcal{A}_i \vDash i$, then there is an ultrafilter $\mathcal{U}$ on I such that $\prod_I \mathcal{A}_i/\mathcal{U} \vDash \Sigma$.*

Proof. For each $\sigma \in \Sigma$, let $\hat{\sigma} = \{i \in I : \sigma \in i\}$. Then the set $\{\hat{\sigma} : \sigma \in \Sigma\}$ has finite intersection property- $\{\sigma_0, \sigma_1, ..., \sigma_n\} \in \hat{\sigma}_0 \cap \hat{\sigma}_1 \cap ... \cap \hat{\sigma}_n \neq \emptyset$, thus by Theorem 3.2 can be expanded to an ultrafilter $\mathcal{U}$ on I. Since $\hat{\sigma} \in \mathcal{U}$ for each $\sigma \in \Sigma$, we have $\hat{\sigma} \subseteq \{i : \mathcal{A}_i \vDash \sigma\} \in \mathcal{U}$. Then from Theorem 3.16, $\prod_I \mathcal{A}_i/\mathcal{U} \vDash \Sigma$. □

Corollary 3.19.[7] *A set of $\mathcal{L}$-structures is an elementary class if and only if it is closed under elementary equivalence and ultraproducts.*

Proof. An elementary class is clearly closed under elementary equivalence and ultraproducts.
Conversely, if a set of $\mathcal{L}$-structures $\mathcal{K}$ is closed under ultraproducts and elementary equivalence, let T be the $\mathcal{L}$-theory of sentences satisfied by all structures in $\mathcal{K}$. If $\mathcal{M}$ is any model of T and $\mathrm{Th}(\mathcal{M})$ is the full theory of $\mathcal{M}$, then every element i in $I = \{X \subseteq \mathrm{Th}(\mathcal{M}) : X \text{ is finite}\}$ is satisfied by some element $\mathcal{A}_i \in \mathcal{K}$ (otherwise $\neg \bigwedge i$ would be satisfied by every element in $\mathcal{K}$ hence $\neg \bigwedge i \in T$ but $\mathcal{M} \nvDash \neg \bigwedge i$, contradiction.) Then by Corollary 3.18 we can construct an ultraproduct $\prod_I \mathcal{A}_i/\mathcal{U}$ which must satisfy $\mathrm{Th}(\mathcal{M})$ hence $\prod_I \mathcal{A}_i/\mathcal{U} \equiv \mathcal{M}$. But since $\mathcal{K}$ is closed under ultraproducts and elementary equivalence, we have $\prod_I \mathcal{A}_i/\mathcal{U} \in \mathcal{K}$ and thus $\mathcal{M} \in \mathcal{K}$. Hence $\mathcal{K}$ is the elementary class axiomatized by T. □

Just to put into perspective how compact[10] the proof of the Compactness theorem in Corollary 3.18 is compared to Corollary 1.17, Henkin's proof of Theorem 1.16 involves adding constant-symbols and special function-symbols to expand the language and theory so that every existential statement is witnessed, expanding the theory to a maximal one, showing it is satisfiable and then translating back to the original theory, the method is called *Henkin constructions* (See [12] for the detailed proof). The expanded theory is called a *Skolemization*, and satisfies the following properties- every

[9] Also you are lying if you say anything with *ultra* in its name is not cool.
[10] Not sorry

model of the original theory can be expanded into a model of the Skolemization, and for every formula $\phi(\bar{x}, y)$ there is a term t in the expanded language such that $\forall \bar{x}(\, \exists y\, \phi(\bar{x}, y) \rightarrow \phi(\bar{x}, t(\bar{x}))\,)$.

3.5 Non-standard Universes

Henkin's method of Skolemization essentially helps us pass from $\forall x\, \exists y \phi(x, y)$ to $\exists F\, \forall x\, \phi(x, F(x))$, i.e. switching the order of quantifiers. Ultraproducts help us with something similar, but instead of adding new functions we add new elements.
For $\mathcal{L} = \{<, +, \cdot, 0, 1\}$ consider the theory

$$\mathrm{Th}(\mathbb{R}) \cup \{\exists x,\ 0 < x < \frac{1}{\underbrace{1 + 1 + ... + 1}_{i \text{ times}}} : i = 0, 1, 2, ...\}.$$

This is finitely satisfiable by $\mathbb{R}$, hence by the compactness theorem it has a model $\mathbb{R}^*$, called the set of *non-standard real numbers.* In addition to all first order properties of $\mathbb{R}$, this set also satisfies sentences like $\exists x\, \forall y,\ x > y$ and $\exists x\, \forall y > 0,\ 0 < x < y$. These properties assert the existence of arbitrarily large and small numbers, and helped Abraham Robinson formalize the ancient but more intuitive approach to calculus through *infinitesimals.*[14]

3.5.1 Hyperreal Numbers[13]

Compactness theorem asserts the existence of a non-standard set of real numbers, ultraproducts help us construct one. Cauchy defined the real numbers[11] as equivalence classes on $\mathbb{Q}^{\mathbb{N}}$. In a similar fashion we define the *hyperreal numbers* $(\mathbb{R}, \mathbb{Z}, +, \cdot, 0, 1)^*$ from $(\mathbb{R}, \mathbb{Z}, +, \cdot, 0, 1)$ to have the universe $\mathbb{R}^* = \mathbb{R}^{\mathbb{N}}/\mathscr{U}$, for some fixed non-principle ultrafilter $\mathscr{U}$ on $\mathbb{N}$. It follows immediately from the transfer principle that $\mathbb{R}^*$ is a field, since all field axioms are expressible as first order sentences.

Henceforth we only write $\mathbb{R}$ and $\mathbb{R}^*$ for the structures. Since the structure specifies all integers (by assigning constant-symbols in the vocabulary), the $*$-transform immediately gives us the *hyperintegers* $\mathbb{Z}^*$ and the *hypernaturals* $\mathbb{N}^*$. Since the existence of multiplicative inverses can be written in first order, we also have the set of *hyperrationals* $\mathbb{Q}^*$.

[11] See Cauchy sequences

By identifying $r \in \mathbb{R}$ with $(r, r, r, ...)/\mathcal{U} = r^* \in \mathbb{R}^*$, we see that the real numbers are a subfield of hyperreals. In fact, $\mathbb{R}$ is a proper subfield, since the hyperreal $(1, 2, 3, ...)/\mathcal{U}$ does not coincide with any real number. Since there are $(2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0}$ sequences in $\mathbb{R}^{\mathbb{N}}$, we see that $\mathbb{R}^*$ has the same cardinality as $\mathbb{R}$.

Functions on $\mathbb{R}$ extend naturally to functions on $\mathbb{R}^*$, by assigning

$$f((a_0, a_1, a_2, ...)/\mathcal{U}) = (f(a_0), f(a_1), f(a_2), ...)/\mathcal{U}.$$

We thus have a well defined notion of the absolute value $|x|$ of any hyperreal x. Writing $\omega = (1, 2, 3, ...)/\mathcal{U}$, we see that $|\omega| > |x^*|$ for every $x \in \mathbb{R}$. Such numbers are called *unlimited hyperreals.* Hyperreals that are not unlimited are called *limited*. On the other hand, writing ϵ for the multiplicative inverse $(1, \frac{1}{2}, \frac{1}{3}, ...)/\mathcal{U}$ of ω, it is true that $0 < |\epsilon| < |x^*|$ for every non-zero $x \in \mathbb{R}$. These hyperreals are called *infinitesimals.*

To illustrate how the transfer principle does not hold for logics stronger than first order, we show how the key second-order property of real numbers (the least upper bound property (2.1)) is false in the hyperreals: $\mathbb{R}$ is a non-empty bounded subset of $\mathbb{R}^*$ without a supremum- for if s is such that $s \geq x^*$ for every $x \in \mathbb{R}$, then s must be unlimited- but then $s - 1^* < s$ is also unlimited and hence there is no least upper bound.

Definition 3.20. Two hyperreal numbers x and y are *close* (written $x \simeq y$) if their difference is infinitesimal or zero.

Closeness is an equivalence relation- the equivalence class of x is termed the *halo* of x, written $\mathrm{halo}(x)$. We write $x \simeq 0$ to denote x is infinitesimal or zero.

Theorem 3.21. *There is exactly one real number in the halo of any limited hyperreal x. This is called the standard part of x, written* $\mathrm{st}(x)$.

Proof. For a hyperreal x, consider the set $A_x = \{r \in \mathbb{R} : \ r^* \leq x\}$. Since x is limited, A_x is a non-empty subset of $\mathbb{R}$ that is bounded above, hence has a least upper bound α. Note we have $x \geq \alpha^*$, i.e. $x - \alpha^* \geq 0$. For every positive real number r, we must have $(\alpha + r)^* \geq x$, which implies $0 \leq x - \alpha^* \leq r^*$, or $x - \alpha^* \simeq 0$. Thus there is atleast one real number in the halo of every limited hyperreal. Suppose α and β are real numbers in the halo of x, then we have $x \simeq \alpha^* \simeq \beta^*$, or $\alpha^* - \beta^* \simeq 0$. But this is possible only if $\alpha = \beta$, hence the standard part of every limited hyperreal is unique. $\square$

Remark. At this point we shall stop writing r and r^* to distinguish between r as a real and a hyperreal number respectively, writing just r for both. However, the two remain to be distinct entities, and which one we are referring to is usually clear from context.

We proceed to illustrate the appeal of hyperreal numbers when formulating calculus:

3.5.2 Non-standard Calculus

Corollary 3.22.[8] *$\mathbb{R}$ is Dedekind complete (every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$.)*

Proof. Let $\langle a_n \rangle$ be a Cauchy sequence in $\mathbb{R}$, i.e. the terms get arbitrarily close to each other. In particular, there exists a k_ϵ such that

$$\forall m, n \in \mathbb{N},\ m, n > k_\epsilon \to |a_m - a_n| < 1.$$

Using the transfer principle, we can take the $*$-transform of this sentence, i.e. for the hypersequence $\langle a_n \rangle : n \in \mathbb{N}^*$,

$$\forall m, n \in \mathbb{N}^*,\ m, n > k \to |a_m - a_n| < 1.$$

Taking m to be the unlimited non-standard natural number ω, we see that $|a_\omega - a_k| < 1$ i.e. a_ω is limited. Hence there exists a unique real number $a \simeq a_\omega$. Again, using the fact that $\langle a_n \rangle$ is Cauchy, for every $\epsilon > 0$ there exists a $l > 0$ such that

$$\forall m, n \in \mathbb{N},\ m, n > l \to |a_m - a_n| < \frac{\epsilon}{2}.$$

Taking the $*$-transform, we see

$$\forall m, n \in \mathbb{N}^*,\ m, n > l \to |a_m - a_n| < \frac{\epsilon}{2},$$

i.e. $|a_n - a_\omega| < \frac{\epsilon}{2}$ whenever $n > l$, and since $a_\omega \simeq a$ we must have

$$|a_n - a| < |a_n - a_\omega| + |a_\omega - a| \leq \frac{\epsilon}{2} + \text{ infinitesimal} < \epsilon$$

whenever $n > l$. This is the exact condition for the sequence to converge to a, hence we are done. $\square$

The a_ω we defined above is in fact close to the hyperreal $\langle a_n \rangle / \mathcal{U}$. This can be shown as follows: for $a' = \mathrm{st}(a_\omega)$, consider the sequence

$$\langle a_n - a' \rangle = (a_0 - a', a_1 - a', a_2 - a', ...).$$

Since a_ω is the hyperreal limit of the sequence, for every non-zero real ϵ, the sequence is greater than $\frac{\epsilon}{2}$ only in finitely many places, hence $|\langle a_n - a' \rangle / \mathcal{U}| < |\epsilon|$. This implies $\langle a_n \rangle / \mathcal{U} \simeq a' \simeq a_\omega$ as desired. This gives us a neat interpretation of limits- every natural number 'votes' what the limit should be, and the limit of the sequence is whichever real number wins the election.[15]

Definition 3.23. We define the *non-standard limit* of a sequence $\langle a_n \rangle$ of real numbers as

$$\lim_{i \to \infty}{}^{*} a_i = \mathrm{st}(\langle a_n \rangle / \mathcal{U}).$$

The notion of a limit of a function follows immediately in classical calculus: we say

$$\lim_{x \to a}{}^{*} f(x) = L$$

if for every sequence $a_i \to a$ we have $f(a_i) \to L$, or (using the definition of functions on hyperreals), $f(x) \simeq L$ whenever $x \simeq a$. Showing that this is equivalent to the $\epsilon - \delta$ definition is routine calculus and can be found in any real analysis textbook. This leads to the very intuitive definition of continuity:

Definition 3.24. $f(x)$ is continuous at $a \in \mathbb{R}$ if and only if $f(x) \simeq f(a)$ whenever $x \simeq a$.

Can we provide an alternate definition of the limit of a function by considering an ultraproduct $\mathbb{R}^{\mathbb{R}} / \mathcal{U}$ for some non-principle ultrafilter $\mathcal{U}$ on $\mathbb{R}$, and then using a voting interpretation? Will this be equivalent to the classical definition?

Example 3.25. Continuity of polynomials: If $f(x) = ax^n$, $a \in \mathbb{R}$, $n \in \mathbb{N}$, for infinitesimal ϵ we have

$$f(x + \epsilon) \simeq a(x + \epsilon)^n \simeq ax^n + \epsilon \cdot \mathrm{O}(1) \simeq ax^n + \epsilon$$

where $\mathrm{O}(1)$ represents a limited number. Clearly, $f(x) \simeq f(c)$ whenever $x \simeq c$, thus x^n is continuous for $n \in \mathbb{N}$. If f and g are two functions continuous at c, we have

$$f(c + \epsilon) + g(c + \epsilon) \simeq f(c) + \epsilon + g(c) + \epsilon \simeq f(c) + g(c)$$

i.e. sum of two continuous functions is continuous, thus showing polynomials (and in fact all powerseries) are continuous.

Differentiability and derivatives make use of two points on the curve to determine the slope of the tangent. Classically this is done by bringing the two points on a secant line arbitrarily close, but now we can do it with infinitesimals.

Definition 3.26.[8] If a function f is defined at $x \in \mathbb{R}$, then the real number L is the derivative of f at x if and only if for every infinitesimal ϵ, $f(x+\epsilon)$ is defined and

$$\frac{f(x+\epsilon) - f(x)}{\epsilon} \simeq L.$$

Example 3.27. Derivative of sine: for $x \in \mathbb{R}$ and an infinitesimal ϵ,

$$\frac{\sin(x+\epsilon) - \sin x}{\epsilon} = \frac{\sin x \cos \epsilon + \cos x \sin \epsilon - \sin x}{\epsilon} \simeq \cos x$$

since $\sin \epsilon = \epsilon - \frac{\epsilon^3}{3!} + ... \simeq \epsilon$ and $\cos \epsilon = 1 - \frac{\epsilon^2}{2!} + ... \simeq 1$.

Bibliography

[1] A G Thomason (a.g.thomason@dpmms.cam.ac.uk). *Numbers and Sets, Example Sheet 1*. en. 2018-19. URL: http://www.dpmms.cam.ac.uk/study/IA/Numbers%2BSets/2018-2019/examples-NS-18-1.pdf.

[2] A G Thomason (a.g.thomason@dpmms.cam.ac.uk). *Numbers and Sets, Example Sheet 3*. en. 2018-19. URL: http://www.dpmms.cam.ac.uk/study/IA/Numbers%2BSets/2018-2019/examples-NS-18-3.pdf.

[3] *Arrow's Impossibility Theorem and Ultrafilters*. URL: https://www.jyotirmoy.net/posts/2013-10-28-arrow-ultrafilter.html (visited on 08/12/2019).

[4] Athar. *Ramsey's Theorem and Ultrafilters*. en. Apr. 2018. URL: https://abdul-quader.com/2018/04/07/ramseys-theorem-and-ultrafilters/ (visited on 09/11/2019).

[5] Timothy Bays. "Skolem's Paradox". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2014. Metaphysics Research Lab, Stanford University, 2014. URL: https://plato.stanford.edu/archives/win2014/entries/paradox-skolem/ (visited on 09/10/2019).

[6] Luca Bellotti. "Tarski on Logical Notions". en. In: (), p. 23.

[7] C.C. Chang and H.J. Keisler. *Model Theory*. Studies in Logic and the Foundations of Mathematics. Elsevier Science, 1990. ISBN: 978-0-08-088007-5. URL: https://books.google.co.uk/books?id=uiHq0EmaFp0C.

[8] R. Goldblatt. *Lectures on the Hyperreals: An Introduction to Nonstandard Analysis*. Graduate Texts in Mathematics. Springer New York, 1998. ISBN: 9780387984643. URL: https://books.google.co.in/books?id=TII-PX%5C_OdloC.

[9] W. Hodges and S.M.S.W. Hodges. *A Shorter Model Theory*. Cambridge University Press, 1997. ISBN: 978-0-521-58713-6. URL: https://books.google.co.uk/books?id=S6QYeuo4p1EC.

[10] W. Hodges et al. *Model Theory*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1993. ISBN: 978-0-521-30442-9. URL: https://books.google.co.uk/books?id=Rf6GWut4D30C.

[11] John MacFarlane. "Logical Constants". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University, 2017. URL: https://plato.stanford.edu/archives/win2017/entries/logical-constants/ (visited on 09/08/2019).

[12] D. Marker. *Model Theory : An Introduction*. Graduate Texts in Mathematics. Springer New York, 2013. ISBN: 978-1-4757-7732-1. URL: https://books.google.co.uk/books?id=vTIVswEACAAJ.

[13] Diego Andres Bejarano Rayo. "INTRODUCTION TO NON-STANDARD ANALYSIS". en. In: (), p. 17.

[14] A. Robinson. *Non-standard Analysis*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 2016. ISBN: 9781400884223. URL: https://books.google.co.in/books?id=Xh-TDAAAQBAJ.

[15] T Tao. *Ultrafilters, nonstandard analysis, and epsilon management*. en. June 2007. URL: https://terrytao.wordpress.com/2007/06/25/ultrafilters-nonstandard-analysis-and-epsilon-management/ (visited on 08/14/2019).

[16] T Tao. *Ultraproducts as a Bridge Between Discrete and Continuous Analysis*. URL: https://terrytao.wordpress.com/2013/12/07/ultraproducts-as-a-bridge-between-discrete-and-continuous-analysis/#more-7171 (visited on 08/14/2019).

[17] *Upward Löwenheim-Skolem Theorem - ProofWiki*. URL: https://proofwiki.org/wiki/Upward_L%C3%B6wenheim-Skolem_Theorem (visited on 09/10/2019).

[18] Jouko Väänänen. "Lindström's Theorem". en. In: *Universal Logic: An Anthology*. Ed. by Jean-Yves Béziau. Basel: Springer Basel, 2012, pp. 231–235. ISBN: 978-3-0346-0144-3 978-3-0346-0145-0. DOI: 10.1007/978-3-0346-0145-0_19. URL: http://link.springer.com/10.1007/978-3-0346-0145-0_19 (visited on 09/11/2019).

[19] Sebastien Vasey. "MATH 141A: ARROW'S IMPOSSIBILITY THEOREM AND ULTRAFILTERS". en. In: (), p. 7.